MONOTONIC MEASURES IN KNOWLEDGE ENGINEERING: AN APPLICATION IN THE PRODUCTION RULE FORMALISM

by

SUJEET SHENOI

B.Tech., Indian Institute of Technology, Bombay, 1981

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Chemical Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1984

Approved by:

F. S. Lai
Co-Advisor

L. T. Fan
Co-Advisor

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENT

Artificial Intelligence, Measure Theory, and Fuzzy Sets are exciting fields of research. It has been my privilege to work in these areas. I should like to express my sincerest thanks to Dr. L. T. Fan for this opportunity, and also for his invaluable guidance, encouragement and advice. I also wish to thank my Co-Major Advisor, Dr. F. S. Lai for his continued support and guidance during the course of this work. I sincerely thank Dr. R. T. Hartley for introducing me to the field of Artificial Intelligence, and also for consenting to sit on my committee. I also wish to express my thanks to my committee member, Dr. Virgil Wallentine of the Department of Computer Science.

Thanks are due to Kenichi Toguchi of Mitsubishi Kakoki Kaisha Ltd., Tokyo, Japan, for his invaluable suggestions and assistance. I also wish to thank Dr. E. Tazaki of Omron Tateisi Electronics Co., Osaka, Japan, and Dr. H. Tanaka of the University of Osaka Prefecture, Osaka, Japan, for enlightening me on various aspects of Fuzzy Set Theory.

The financial support has been provided by the U.S. Department of Agriculture-Kansas State University Broad Form Cooperative Agreement No. 5340. All experimental work has been conducted at the U.S. Grain Marketing Research Laboratory, Manhattan, Kansas.

Lastly, I wish to express my thanks to Chetan Mehta, Snehal Patel, and Virginia Sylvester for their help during the course of this work.

CHAPTER I

INTRODUCTION

Artificial Intelligence (AI) is the branch of computer science that attempts to have machines emulate intelligent human behavior. This goal is rather formidable, and, until recently, workers in the field had met with limited success. However, research activity directed towards reasoning from knowledge in restricted domains has produced computer programs that often approach, and, at times, even exceed human levels of performance. This has led to a renewed interest in AI (Kinnucan, 1984) which has spawned buzz-phrases such as *Knowledge Engineering* and *Expert Systems*.

The recent successes encountered in AI research are due to shifts in the philosophy and direction of conventional approaches. The older methodologies sought for generalized increases in computational power, and researchers strived to develop techniques that were as general as possible (Minsky and Papert, 1974). Unfortunately, such strategies proved to be hopelessly inefficient for dealing with the sheer combinatorial complexity that was often encountered in real-world problems. The *knowledge-based* or *epistemic* strategy (Feigenbaum, 1977), on the other hand, is a pragmatic approach toward the emulation of intelligent human activity. The approach emphasizes domain-specific problem solving strategies over the older, *weaker* methods. Progress is seen as coming from better ways to express, recognize, and use diverse and particular forms of knowledge. The approach recognizes explicitly the local quality of human expertise, and we are encouraged to attack realistic problems that may have been suitably constrained so that useful solutions are realized. The shift to the knowledge-based approach

in AI has contributed to the fast-growing sub-field of Knowledge Engineering and the development of Expert Systems.

Knowledge is a precious resource, and Knowledge Engineering is concerned with the tasks of extracting, articulating, and computerizing knowledge. In much the same way as electricity is the power source for electrical engineers, knowledge engineers view knowledge as a source of power. They *tinker* with knowledge, and the appliances they create *run* on knowledge.

Expert systems may be considered to be the appliances created by Knowledge Engineering. They are computer programs that embody knowledge and use it to solve real-world problems in specific areas of human activity. These programs use collections of facts, rules of thumb, and other forms of domain-specific know-how, coupled with methods for applying this knowledge, to make inferences. They differ substantially from conventional computer programs because their tasks have no simple algorithmic solutions, and because often they must reason in the presence of incomplete and uncertain information.

Production rules (also known as *condition-action*, or *IF-THEN rules*; see, e.g., Barr and Feigenbaum, 1981) are a popular approach for representing and manipulating domain knowledge in expert systems. *Rule-based*, or *Production Systems* (Davis and King, 1977), as they have come to be called in the jargon of AI, operate by selecting rules, verifying whether the premises or condition parts of the rules are satisfied, noting the results, and applying new rules based on the changed situation. Production rules are natural to human strategies of problem solving and

decision making, and this has contributed to their application in many expert systems.

An expert system that relies on a pattern of rule-directed inference represents an attempt to capture the spirit of human reasoning in a computer program. This is consistent with the goals of AI research. However, human reasoning is characterized by an ability to reason in qualitative and imprecise terms; and this ability must be introduced into the production rule formalism. Several expert systems (see, e.g., MYCIN, Shortliffe, 1976), in their attempts to emulate human strategies, *soften* production rules so that partial satisfaction of their premises are sufficient to lead to certain actions or decisions being made. In a *soft production rule*, the propositions that make up the premise are permitted to take truth values in the interval ranging from complete truth to absolute falsehood. This represents a departure from the domain of conventional two-valued logic to one of multivalued logic. Most theories of approximate reasoning (see, e.g., Zadeh, 1975a, 1975b, 1975c) are founded on a multivalued logic base, and could be justified on the basis of observation of human behavior in the real-world.

It is well-known that human experts introduce considerable subjectivity into their decision making. When performing evaluations, they are inclined to weigh and balance the evidence. From this point of view, it is not sufficient that propositions be allowed to take multivalued truth levels; but it is also necessary that we incorporate methodologies for combining these truth values, so that the evidence is, indeed, weighed and balanced.

The combination of the separate pieces of evidence, provided by individual propositions contained in premises of production rules, is distinctly non-linear, and requires meta-level descriptions of knowledge. It is a fact that some propositions are more important than others, and would, therefore, carry greater weights in the evaluation of a premise. This deeper information concerning the relative weights of propositions might be a significant feature of domain-specific know-how. The essential characteristic of an expert, perhaps, is that he possesses accurate conceptions of these weights. Doubtless, these *a priori* notions usher in the subjectivity that sways his evaluations. As knowledge engineers, we must look toward ways to express, represent, and use this meta-level knowledge in synthetic models of human reasoning. In doing so, we are able to introduce the non-linear and subjective aspects of human expertise into the framework of mechanistic decision making.

The objective of the present work is to develop a methodology for the combination of evidence in the production rule formalism. This methodology must model effectively the subjectivity that is a feature of human evaluative strategies, and will be applied to the problem of classification of milled rice grain.

AN OVERVIEW OF THE METHODOLOGY

Any methodology that is devised for the combination of evidence must be founded on a basic human trait. *Monotonicity* is one such fundamental concept that appears to play a major role in human evaluative strategies. The principle of monotonicity is illustrated by the adage:

Given more, we feel at least as good, or even better.

This principle accurately reflects human behavior, and has special relevance in modeling the subjective combination of evidence.

Very often, in real-life, a single piece of evidence is not sufficient to force an evaluation in the direction of truth or falsehood. On the other hand, as additional pieces of evidence are obtained, the total weight of the body of evidence pointing to the conclusion becomes greater, the picture begins to clear, and the evaluation becomes more certain, or, at least, remains the same. This process conforms with the principle of monotonicity.

The present methodology focuses on the combination of evidence in the production rule formalism. The premise of a rule comprised of AND-connected propositions is written as a set. Each proposition is an element of the premise set, and is considered to represent a specific piece of evidence that points toward the action. Thus, the evaluation of the premise essentially involves the combination of the distinct pieces of evidence provided by the individual propositions. Since the premise is expressed as a set, it is convenient to employ measures of

subsets of this set to quantify the relative weights that groups of propositions, or bodies of evidence, carry in an evaluation. These weights are the *a priori* notions that are used by human beings when they weigh and balance the evidence. The combination of the measures models the combination of the evidence provided by individual propositions. Human combination of evidence is performed monotonically. The measures, therefore, must obey the same principle. Specifically, *fuzzy measures* (Sugeno, 1974) are employed in the present work to quantify the relative importances that groups of propositions carry. These measures follow the principle of monotonicity, and for this reason, we prefer to call them *monotonic measures*. This term has been employed throughout this thesis.

If the propositions comprising the premise are allowed to adopt just one of two truth values - *true* and *false*, the combination of measures is sufficient to model the combination of evidence. However, the present methodology permits multivalued truth levels, and it is necessary to combine these truth values with the relative weights in a premise evaluation. The *Sugeno Integral*, also known as the *Fuzzy Integral* (Sugeno, 1974), defined on monotonic measure space, unites these two quantities. The result is a *mean* or weighted premise evaluation that is also monotonic, and possesses excellent intuitive justification.

As will be seen in later chapters, several significant advantages are offered by the application of monotonic measure theory in the framework of production rules.

i) The methodology employing monotonic measures provides a

convenient foundation for expressing, representing, and coping with the subjectivity that is the hallmark of human evaluative strategies.

ii) It offers a viable framework for the representation and treatment of ignorance, and the conservatism that is seen in evaluations made in its presence.

iii) The Sugeno Integral is simply an extension of the minimum operator that is conventionally used to evaluate premises consisting of AND-connected propositions. In fact, it will be seen that the Sugeno Integral reduces to the minimum operator in the absence of meta-level knowledge about the relative weights of propositions.

iv) The methodology may be extended to admit multilevel reasoning without any loss in generality.

The guidelines imposed by Knowledge Engineering exhort us to search for better ways to express, recognize, and use diverse and particular forms of knowledge. This is our intention, and the advantages gain significance when viewed in this light.

ORGANIZATION OF THE THESIS

In this thesis, we employ elements of monotonic measure theory to model the human subjective combination of evidence in the production rule formalism. The current emphasis on the proper representation and use of knowledge has motivated the development of the present methodology, and in Chapter II, we review the significance of the knowledge-based approach in AI. Expert systems are a natural consequence of this approach, and we proceed to examine some of the more important issues in expert systems research.

Chapter III provides the theoretical foundation for the methodology. In order to gain an insight into the controversy surrounding non-additive probability, we begin by tracing the historical conceptions of probability. We go on to examine monotonic measures, and finally, focus on the Sugeno Integral, a functional defined on monotonic measure space.

Chapter IV forms the core of the present work. We start by examining production systems in considerable detail. The concept of monotonic measures is introduced into the production rule formalism, and a methodology based on the Sugeno Integral is proposed for the evaluation of premises of production rules. The methodology is shown to offer a convenient framework for the treatment of ignorance, and is subsequently extended to admit multilevel reasoning.

In Chapter V, the methodology is applied to the problem of classification of rice grain. An attempt is made to follow the visual approach that an expert grain inspector would adopt. Essentially, a prototype

from each class is defined by a unique production rule, similar in form to a discriminant function employed in classical pattern recognition theory.

The conclusions and recommendations for future work are summarized in Chapter VI.

REFERENCES

1. Barr, A., and Feigenbaum, E. A., "The Handbook of Artificial Intelligence: Vol. 1", William Kaufman Inc., Los Altos, Calif. (1981).

2. Davis, R., and King, J. J., "An Overview of Production Systems", In Elcock, E., and Michie, D., (Eds.), "Machine Intelligence 8", Ellis Horwood, Chichester, U.K., 300-332 (1977).

3. Feigenbaum, E. A., "The Art of Artificial Intelligence: I. Themes and Case Studies of Knowledge Engineering", Int. Joint Conf. Artif. Intell., 5, 1014-1029 (1977).

4. Kinnucan, P., "Computers that think like experts", High Tech., 30-42 (Jan., 1984).

5. Minsky, M., and Papert, S., "Artificial Intelligence", Condon Lectures, Oregon State System for Higher Education, Eugene, Oregon (1974).

6. Shortliffe, E. H., "Computer-based Medical Computations: MYCIN", American Elsevier, New York (1976).

7. Sugeno, M., "Theory of Fuzzy Integrals and its Applications", Ph.D. Thesis, Tokyo Institute of Technology, Tokyo (1974).

8. Zadeh, L. A., "The concept of a linguistic variable and its application to approximate reasoning, Part I", Inf. Sc., 8, 199-249 (1975a).

9. Zadeh, L. A., "The concept of a linguistic variable and its application to approximate reasoning, Part II", Inf. Sc., 8, 301-357 (1975b).

10. Zadeh, L. A., "The concept of a linguistic variable and its application to approximate reasoning, Part III", Inf. Sc., 9, 43-80 (1975c).

CHAPTER II

ARTIFICIAL INTELLIGENCE, KNOWLEDGE ENGINEERING, AND EXPERT SYSTEMS

Ever since Charles Babbage conceived his *Difference* and *Analytic Engines* in the mid-nineteenth century, mankind has devoted considerable effort toward machine-based creativity. The advent of the first digital computers in the early 1950s revolutionized this effort, and attempts to develop *thinking machines*, that could emulate intelligent human behavior, seemed destined for success. *Artificial Intelligence* (*AI*) is the discipline that is devoted to developing and applying computational approaches to intelligent behavior. However, researchers in the field always fell short of their goal, the creation of a genuine thinking machine.

In the mid-1960s, AI underwent a shift to a *knowledge-based* paradigm. The new approach emphasizes the power of knowledge, and has led to the creation of a new sub-field called *Knowledge Engineering*. Knowledge Engineering is the technology that promises to make knowledge a valuable commodity, and, in recent years, research in this area has had many major successes. Perhaps, the most noteworthy of these has been the construction of *Expert Systems*. Modeled on human experts, these programs are designed to represent and apply domain-specific knowhow in solving practical problems. Several conventional systems have been evaluated as performing at or above the level of human experts. As a result, interest in expert systems has exploded in industry and government.

In this chapter, we review the significance of the shift in emphasis of AI research from the older, *weaker* methods to the knowledge-based approach and Knowledge Engineering. Expert systems are a natural

development of this approach, and we proceed to examine some of the more important issues in expert systems research.

THE KNOWLEDGE-BASED APPROACH IN ARTIFICIAL INTELLIGENCE

The realization that digital computers are not just fast adding machines, but instead, are potentially capable of being programmed to exhibit human-like intelligence has sparked serious interest in Artificial Intelligence (AI). AI is perceived as the computational study of intelligence, and researchers in the field attempt to develop computational models of intelligent behavior, including both its cognitive and perceptual aspects [see, e.g., Barr and Feigenbaum, 1981, 1982.]. In practical terms, this reduces to the development of computer programs that can solve problems normally thought to require human intelligence.

The earliest years (the late 1950s and early 1960s) saw attempts to solve problems that had a distinctive non-numerical flavor. Computers were programmed to play games, compose music, solve puzzles, and even, devise and prove theorems in mathematics and symbolic logic. On the theoretical side, the important techniques that emerged emphasized the symbolic aspects of problem solving. Researchers looked for structures for representing symbolic information, methods for manipulating these structures and heuristics for searching through them. While the results obtained during this period supported the possibility of machine intelligence, they could not provide a basis for solving complex practical problems.

Goldstein and Papert (1977) discern important reasons for this failure. The early period confined AI to the domain of heuristic search,

that is, the study of procedural techniques for exploring state spaces too large to be explored exhaustively. This was due to the feeling that a relatively small number of powerful general mechanisms would be sufficient to generate intelligent behavior, and manifested itself in the *power-based approach.* Minsky and Papert (1974) characterize this point of view.

"The *power strategy* seeks a generalized increase in computational power. It may look toward new kinds of computers (*parallel* or *fuzzy* or *associative* or whatever) or it may look towards extensions of deductive generality, or information retreival, or search algorithms - things like better *resolution* methods, better methods for exploring trees and nets, hash-coded triplets, etc.. In each case the improvement sought is intended to be *uniform* - independent of the particular data base."

Experience showed that programs that relied on uniform search or logistic techniques which were problem-independent proved to be hopelessly inefficient for handling the sheer combinatorial complexity that was often encountered. Additionally, according to Duda and Shortliffe (1983). the general techniques were found to be inadequate when confronted with imprecisely stated *problems*, uncertain *facts*, and unreliable *axioms*.

Today, the most fundamental problem in AI is not the identification of a few powerful techniques. Instead, as suggested by Goldstein and Papert (1977), is the question of "how to represent large amounts of knowledge in a fashion that permits their effective use and interaction". It is realized that there are diverse kinds of knowledge, and the problem-solver, whether man or machine, must know how to process the knowledge it has. For this reason, it is imperative that the general

techniques of the older approach be supplemented with "domain-specific pragmatic knowhow". Thus, there has been a shift from the *power strategy* to what is termed as a *knowledge-based* or *epistemic approach.* In the words of Minsky and Papert (1974),

> "The *knowledge strategy* sees progress as coming from better ways to express, recognize, and use diverse and particular forms of knowledge. This theory sees the problem as epistemological rather than as a matter of computational power or mathematical generality. It supposes, for example, that when a scientist solves a new problem, he engages a highly organized structure of especially appropriate facts, models, analogies, planning mechanisms, self-discipline procedures, etc.. To be sure, he also engages in *general* problem solving schemata but it is by no means obvious that very smart people are that way directly because of the superior power of their general methods - as compared with average people. Indirectly, perhaps, but that is another matter: a very intelligent person might be that way because of specific local features of his knowledge-organizing knowledge rather than because of global qualities of his thinking which, except for the effects of his self-applied knowledge, might be little different from a child's."

The knowledge-based approach serves to identify AI as a procedural theory of knowledge. The view is that the process of intelligence is determined by the knowledge held by the subject, and the approach stresses an understanding of the operations and data structures involved. We can discern two key procedural concerns. Knowledge within a specific domain must be represented so that it can be used efficiently. Comprehension, transformations, and results must occur within a reasonable length of time. So the first concern is to identify and formalize domain-specific knowledge. Most intellectual activity involves the interaction of knowledge from different domains. Hence, the second, and essential concern, is to construct frameworks so that diverse kinds of knowledge

can successfully interact. In this view, AI embraces attempts to structure knowledge into procedural systems that can solve complex real-world problems.

The shift to the knowledge-based approach, while delivering AI from its initial forays into *toyland*, has contributed to a fast-developing sub-field called *Knowledge Engineering*. Feigenbaum (1977) has defined this activity as

"...the art of bringing the principles and tools of AI research to bear on difficult applications problems requiring experts' knowledge for their solution. The technical issues of acquiring this knowledge, representing it, and using it appropriately to construct and explain lines-of reasoning, are important problems in the design of knowledge-based systems. ... It is the art of building complex computer programs that represent and reason with knowledge of the world."

Feigenbaum's definition of Knowledge Engineering is a prescriptive guide for the construction of *Expert Systems*. An expert system is a computer program that can help solve complex, real-world problems in a specific area of human expertise. The development of expert systems is the result of the shift to the knowledge-based approach, and these programs are characterized by their use of large bodies of domain-specific knowledge. Human experts normally possess extensive knowledge about a narrow class of problems. It is this feature that makes it feasible to provide a computer with sufficient knowledge so that it could serve as a consultant for decision making.

The field of expert systems is perhaps the most active area of applied research in AI. Several factors have motivated this development.

In many areas of human expertise, problems are very complex, and this often results in large solution spaces. A large solution space renders it difficult for a human being to locate all possible solutions, or even, to be confident of a particular solution. For a computer, this limitation is not too severe, and it can effectively search a large solution space if it is provided with a proper conceptual methodology. Often, the same methodology can be used to search an even larger solution space with no significant increase in computational time. Additionally, in domains such as medical diagnosis, problem solving by computer also ensures that remote possibilities are not overlooked. Otherwise, a potentially disastrous situation is likely. An expert system could, therefore, provide reliable and thorough services, more rapidly, and perhaps, at a reduced cost.

Another motivation is due to the fact that some tasks are too routine for a human being to perform repeatedly. It is a good idea to delegate such tasks to a computer. In these scenarios, the human assumes the role of a supervisor; to man is allotted the task he does best - *thinking*.

IMPORTANT ISSUES IN EXPERT SYSTEMS RESEARCH

The simplest and most successful expert systems are classification programs. These systems, which are designed to be used in a well-defined context, weigh and balance pieces of evidence for a given manifestation to decide how it should be classified. A number of consultation systems which are used as aids in medical decision-making fall in this category. The expert system, MYCIN (Shortliffe, 1976), is designed to provide consultative advice on the diagnosis and therapy for microbial infectious diseases. CASNET (Weiss _et al_., 1977) aids in the assessment and treatment of Glaucoma. PUFF (Kunz _et al_., 1978) is being used to analyze pulmonary function tests. Other important medical applications systems are INTERNIST (Pople, 1975), for internal medical diagnosis; the Digitalis Therapy Advisor (Silverman, 1975); and EXPERT (Weiss and Kulikowski, 1979), a general facility that helps investigator build medical consultative models in Rheumatology, Opthalmology, and Endocrinology.

Expert systems have not been confined to medical diagnosis alone, and DENDRAL (Buchanan and Feigenbaum, 1978), is perhaps the best known system from outside this domain. It is one of the earlier expert systems, and predicts the chemical structure of unknown compounds by analyzing mass spectral patterns. Also well-known is PROSPECTOR (Duda _et al_., 1979). PROSPECTOR assists geologists in hard-rock mineral exploration. Some of the other important expert systems are R1 [or XCON (McDermott, 1982)], which is being used by Digital Equipment Corporation (DEC) to configure

VAX computers; HASP and SIAP (Nii et al., 1982), which use information about vessels and the sea, and expertise about signal interpretation to analyze signals from ocean sensors; and DELTA (Bonissone, 1983), which is used by General Electric Corporation for troubleshooting diesel electric locomotives.

Many of these systems are considered to have achieved performances at the expert level. This success is, in part, due to the fact that much of the experts' knowledge in these domains concern specific pieces of information, which has made it easier to identify and filter the necessary knowledge. In contrast, Duda and Shortliffe (1983) opine that it is much more difficult to develop expert systems that have a more *synthetic* character, such as those that concern planning or require *de novo* generation of solutions. There are other basic problems that have been holding back the wide proliferation of expert systems. We proceed to examine some of the more important issues in expert systems research.

Knowledge Acquisition

Building an expert system requires the transfer of expertise to a computer program. The identification and representation of this knowledge is complex and presents many problems. Experts often have difficulty expressing their knowledge in the knowledge representation formalism that is being used. Currently, the only successful method of knowledge transfer is through a computer scientist intermediary.

Attempts to construct knowledge bases often disclose inadequacies

in our understanding of the subject domain. Human beings also tend to forget, or to simplify details about their expertise. Additionally, the domains themselves develop rapidly with time, and it is necessary for the system to augment its knowledge at a later date.

If an expert system is to perform as well as human experts, it should be able to learn as they do. Current research is geared towards the development of learning systems as a means of knowledge acquisition. A potential solution is to allow the expert to teach the system directly. It is realized that learning is not just the accumulation of new facts, but instead, involves the interaction of old and new knowledge. From this point of view, it is important to understand how human experts talk about what they know, and it has been suggested (Duda and Shortliffe, 1983), that, in the design of systems that allow for interactive transfer of expertise, the machine should be able to ask focused queries, and not general questions. The system must also be able to make changes in its knowledge base, and it must do so easily, and in an incremental or modular fashion [see, e.g., TEIRESIAS for EMYCIN systems (Davis, 1976)]. It is, therefore, obvious that proper knowledge representation is also an essential concern in the design of expert systems.

Knowledge Representation

Efficient knowledge representation is the key issue at this point in the development of AI. Most researchers adopt a pragmatic view of knowledge representation. In this view, a knowledge representation

formalism is a combination of data structures and interpretative procedures that if used properly by a program could lead to *knowledgeable behavior*. Duda and Shortliffe (1983) list the roles that a knowledge representation formalism must assume; they are:

i) Faithful representation of the concepts and intentions of the expert.

ii) Allow for effective and correct interpretation by the program.

iii) Support explanations that convey a line of reasoning that a human expert can understand and critique.

iv) Facilitate the process of finding gaps and errors in the knowledge base.

v) Allow the separation of domain knowledge from the interpretation program so that the knowledge base can be enlarged or corrected without the need for reprogramming the interpreter.

The last three properties point toward a single, uniform formalism that is simple and easy to interpret. This methodology has been used successfully in many expert systems. However, recently there has been a trend towards more complex and heterogeneous representation schemes that would allow for faithful representation and effective interpretation (Stefik et al., 1982).

Lately, research in knowledge representation schemes has involved the design of several classes of data structures for storing information. They include *logic*, *production rules*, *semantic networks* and *frames*. The flexibility and precision of mathematical logic make it a useful method

and also promote it as a basis for comparing different representation schemes. Production rules offer a modular and uniform mechanism which has been used successfully in several conventional expert systems. Semantic networks simplify certain deductions (inferences through taxonomic relations) by reflecting them directly in the network. Frames generalize this idea by providing frameworks or structures for organizing knowledge. For an excellent treatment of these knowledge representation schemes, the reader is referred to Barr and Feigenbaum (1981).

Research on expert systems has benefitted from the simplicity of using uniform representation schemes. However, significant penalties are incurred when these formalisms are used in large knowledge bases. Stefik et al. (1982), suggest that future research should look into methods of *tuning* expert systems by making changes in the ways they represent knowledge. This would involve the use of specialized data structures, knowledge compilation schemes, and knowledge transforms for *cognitive economy*. Such a system could automatically improve its performance by changing its internal representation.

Other open problems in knowledge representation include *quantification*, that is, the ability to specify properties of arbitrarily defined sets; the representation of beliefs, degrees of certainty, mass nouns, time and tense information, and processes that consist of sequenced actions taking place over time (Barr and Feigenbaum, 1981).

Reasoning and Inference

The ability to reason is intrinsic to human intelligent behavior, and substantial research has been directed towards the development of reasoning mechanisms for expert systems. The process of reasoning, which must resemble that a human expert would employ, usually involves the creation of hypotheses and their verification by weighing and balancing the different pieces of evidence. In a chain of reasoning, the hypotheses are nested, and hypotheses at one level are successively used to verify hypotheses at a higher level. We can discern two specific methodologies. A system could either start from the goals and reason backwards to the data, or, it could reason forwards from data to goals. Most conventional expert systems employ one of these two strategies. The specific domain of application and the architecture of the system's data base play important roles in the selection of the appropriate strategy.

In MYCIN (Shortliffe, 1976), expert reasoning is represented by condition-action rules, which, while linking patient data to infection hypotheses, also provide estimates of certainty for the links. The reasoning process chains backwards from hypothesized diagnoses (goals). Rules are used to estimate the certainty of conclusions based on the certainty factors of their antecedents, to see if the evidence (data) supports a diagnosis. All possible hypotheses are evaluated, after which, MYCIN matches treatments to all diagnoses that have certainties higher than a predefined threshold value. This is termed as a *goal-driven* or *backward-chaining* inference strategy.

The expert system R1, on the other hand, uses a *data-driven*

mechanism to configure VAX computer systems. The user initially enters all the information about the problem, and the rules *chain forward* to evolve the best possible configuration. This strategy is appropriate because the computer configuration problem can be solved without backtracking and without undoing previous steps (Gevarter, 1983).

A new level of complexity is introduced when the expert system must be designed so that it is able to reason in the presence of uncertainty. Conventional systems cope with unreliable or incomplete data in a variety of ways. One of the earliest approaches has been incorporated in MYCIN. This approach employs a model of approximate implication involving a calculus of certainty factors to indicate the strengths of heuristic rules. Although MYCIN has been demonstrated as having expert skills in clinical tests (Yu, Buchanan et al., 1979), several researchers opine that the methodology may be *ad hoc*, at best, since the operational meaning of the computed values is not always clear (Stefik et al., 1982; Duda and Shortliffe, 1983). Other expert systems use methods based on statistical theories. For example, PROSPECTOR assigns probabilities to conclusions using a form of Bayes' rule to update probabilities as more information is obtained. The major drawback in using Bayes' posterior probabilities is that a large number of observations is needed to determine them. This is often not possible, and as a result, the approach may not be statistically valid. Duda et al. (1976) suggest an alternative approach based on subjective estimates of prior probabilities. Other methods for increasing reliability by combining evidence are based on non-additive monotonic

measures (see, e.g., Zadeh, 1978; Shafer, 1976; and Martin-Clouaire and Prade, 1983); and exact approaches using non-monotonic correction rules (Stefik, 1978). All the approaches mentioned require the use of *meta-level descriptions of knowledge*, but a general methodology is lacking. Duda and Shortliffe (1983) are of the opinion that *Possibility Theory* (Zadeh, 1978), or the *Dempster-Shafer Theory of Evidence* (see, e.g., Shafer, 1976), could be used as the basis for formal treatments of imprecision and uncertainty.

A human expert incorporates substantial subjectivity in his decision-making, and this feature has yet to be introduced successfully in expert systems. Future research concerning reasoning mechanisms is also expected to involve reasoning in the presence of ignorance, the ability of a program to recognize the limits of its knowledge, and, when required, engage in cautious guesswork.

Explanation

Explanation of the program's line of reasoning is an important factor for the acceptance of an expert system. Like a human expert, the system must be able to provide explanations about its behavior. A user may need clarification or reassurance about the program's output. An explanation facility contributes to the *transparency* of the reasoning process, and is an essential feature of medical consultation systems. Furthermore, causal explanations also help in the debugging process. Here, the human expert could use the explanations provided to locate the causes of error. The specific knowledge representation formalism

used is an important consideration, and in certain cases, it may be necessary to augment empirical knowledge with causal links to represent functional behavior.

Another feature of human experts is that their explanations are adjusted to satisfy the perceived needs of the user. An expert system should, therefore, maintain a user profile. It must assess what the user does and does not know, and what he is trying to accomplish. This feature is especially important for the instructional use of expert systems (see, e.g., MACSYMA Advisor: Genesereth, 1979). Presently, however, research in this direction is in its infancy, and these features often create more problems than they solve (Duda and Shortliffe, 1983).

Justification and Validation

These are important factors that must be considered before an expert system is deemed fit for general use. A panel of experts must assess the accuracy, reliability, and utility of the system. This would involve examining whether the knowledge representation formalism effectively captures the experts' conceptions of the problem, and whether the associations represented in the system's data base are justified. In order to facilitate justification, a useful design methodology would relate the reasoning steps to deeper causal models using split-level representations.

In some instances, a consensus opinion of the architecture of an expert system may be sufficient to validate it. However, medical consultation systems are also put through years of clinical tests to

verify their performances. CASNET and MYCIN have been rated in experimental evaluations as performing at human-expert levels in their domains (see, e.g., Yu, Buchanan et al., 1979; Yu, Fagan et al., 1979).

The success of expert systems has led to several ethical and sociological problems, especially concerning their use in sensitive areas such as medical diagnosis and nuclear defence. The idea of a computer going beserk has been the subject of several thrillers, such as the recent movie, *WarGames*. A word of caution is in place. Expert systems must be designed to provide advice only if and when the need arises, and under no circumstances should they be allowed to usurp the roles of physicians or missile defence strategists.

For further details concerning expert systems, the interested reader is referred to Hayes-Roth et al. (1983), Barr and Feigenbaum (1981, 1982), Gevarter (1983), and Kinnucan (1984).

CONCLUDING REMARKS

The knowledge-based strategy in AI lays stress on the expression, recognition, and use of diverse and particular forms of knowledge. It acknowledges the local quality of human expertise, and has led to the creation of a new sub-field called Knowledge Engineering. Knowledge Engineering concerns itself with the technical issues of acquiring, representing, and using knowledge in constructing complex computer programs - Expert Systems.

The aim in expert systems research is to develop programs that are able to provide expert-level advice in various domains of human activity. This does not mean that an expert system is viable only if it duplicates intelligent behavior in all its aspects. At present, most expert systems are not able to converse in idiomatic natural language, nor can they perceive evidence directly and learn from experience. Often, they are not able to reason at the expert-level, or even possess elements of common-sense knowledge. We must adopt a pragmatic view and look toward the utility of expert systems. At this point in time, the intention has been to solve realistic problems that have been constrained so that useful solutions are obtained. However, the development of high-performance computer programs is not the only contribution of expert systems research. An equally important contribution is the systematization and codification of domain-specific knowledge. This often leads to new insights, and contributes towards progress within the domains themselves.

REFERENCES

1. Barr, A., and Feigenbaum, E. A., "The Handbook of Artificial Intelligence: Vol. 1", William Kaufman Inc., Los Altos, Calif. (1981).

2. Barr, A., and Feigenbaum, E. A., "The Handbook of Artificial Intelligence: Vol. 2", William Kaufman Inc., Los Altos, Calif. (1982).

3. Bonissone, P. P., "DELTA: An expert system for troubleshooting diesel electric locomotives", paper presented at NAFIP-II Workshop, Schenectady, New York (1983).

4. Buchanan, B. G., and Feigenbaum, E. A., "DENDRAL and meta-DENDRAL: their applications dimension", J. Artif. Intell., 11, 5-24 (1978).

5. Cohen, P. R., and Feigenbaum, E. A., "The Handbook of Artificial Intelligence: Vol. 3", William Kaufman Inc., Los Altos, Calif. (1982).

6. Davis, R., "Applications of meta-level knowledge to the construction, maintenance, and use of large knowledge bases", Memo AIM-283, A.I. Laboratory; and Rep. No. STAN-CS-76-552, Computer Science Dept., Stanford University, Doctoral Dissertation (1976). Reprinted in Davis, R., and Lenat, D. (Eds.), "Knowledge-based Systems in Artificial Intelligence", McGraw-Hill, New York, 227-490 (1982).

7. Duda, R. O., Hart, P. E., and Nilsson, N. J., "Subjective Bayesian methods for rule-based inference systems", SRI International, Artificial Intelligence Center Technical Note (124) (Jan., 1976).

8. Duda, R. O., Gaschnig, J., and Hart, P. E., "Model design in the PROSPECTOR consultant system for mineral exploration", in Michie, D. (Ed.), "Expert Systems in the Micro-electronic Age", Edinburgh Univ. Press, Edinburgh, U.K., 153-167 (1979).

9. Duda, R. O., and Shortliffe, E. H., "Expert systems research", Science, 220, No. 4594, 261-268 (1983).

10. Feigenbaum, E. A., "The Art of Artificial Intelligence: I. Themes and Case Studies of Knowledge Engineering", Int. Joint Conf. Artif. Intell., 5, 1014-1029 (1977).

11. Cenesereth, M. R., "The role of plans in automated consultation systems", Int. Joint Conf. Artif. Intell., 6, 311-319 (1979).

12. Gevarter, W. B., "Expert systems: limited but powerful", IEEE Spectrum, 20, No. 8, 39-45 (1983).

13. Goldstein, I., and Papert, S., "Artificial Intelligence, Language, and the Study of Knowledge", Cog. Sci. J., 1, 84-121 (1977).

14. Hayes-Roth, F., Waterman, D. A., and Lenat, D. B., "Building Expert Systems", Addison-Wesley, Reading, Mass. (1983).

15. Kinnucan, P., "Computers that think like experts", High Tech., 30-42 (Jan., 1984).

16. Kunz, J., et al., "A physiological rule-based system for interpreting pulmonary function test results", Heuristic Programming Project Rep. No. HPP-78-19, Computer Science Dept., Stanford University (1978).

17. Martin-Clouaire, R., and Prade, H., "On the problem of representation of uncertainty in expert systems", paper presented at NAFIP-II Workshop, Schenectady, New York (1983).

18. McDermott, J., "R1, A rule-based configurer of computer systems", Artif. Intell., 19, 38-88 (1982).

19. Minsky, M., and Papert, S., "Artificial Intelligence", Condon Lectures, Oregon State System for Higher Education, Eugene, Oregon (1974).

20. Nii, H. P., Feigenbaum, E. A., Anton, J. J., and Rockmore, A. J., Artif. Intell. Magazine, 3, 23 (1982).

21. Pople, H., "The formation of composite hypotheses in diagnostic problem solving - an exercise in synthetic reasoning", Int. Joint Conf. Artif. Intell., 5, 1030-1037 (1977).

22. Shafer, G., "A Mathematical Theory of Evidence", Princeton Univ. Press, Princeton, N. J. (1976).

23. Shortliffe, E. H., "Computer-based Medical Computations: MYCIN", American Elsevier, New York (1976).

24. Silverman, H., "A Digitalis Therapy Advisor", Rep. No. TR-143, MAC Project, Computer Science Dept., Massachusetts Institute of Technology (1975).

25. Stefik, M., "Inferring DNA structures from segmentation data", Artif. Intell., 11, 085-114 (1981).

26. Stefik, M., et al., "The organization of expert systems, a tutorial", Artif. Intell., 18, 135-173 (1982).

27. Weiss, S. M., and Kulikowski, C. A., "EXPERT: a system for developing consultation nodes", Int. Joint Conf. Artif. Intell., 6, 942-947 (1979).

28. Weiss, S. M., Kulikowski, C. A., and Safir, A., "A model consultation system for the long-term management of glaucoma", Int. Joint Conf. Artif. Intell., 5, 826-832 (1977).

29. Yu, V. L., Buchanan, B. G., et al., "Evaluating the performance of a computer-based consultant", Computer Programs in Biomedicine, 9, 95-102 (1979).

30. Yu, V. L., Fagan, L. H., et al., "Antimicrobial selection by computer - a blinded evaluation by infectious disease experts", J. Amer. Med. Assoc., 242, 1279-1282 (1979).

31. Zadeh, L. A., "Fuzzy sets as a basis for a theory of possibility", Fuzzy Sets and Systems, 1, 3-28 (1978).

CHAPTER III

ON MONOTONIC MEASURES

The past few years have seen a rush to design expert systems in various areas of human activity. In the simplest sense, an expert system is a computer program which contains relevant information and the techniques necessary to manipulate this information, so that, for all practical purposes, it could function at the same level of competence as a human expert in the specific domain. An essential feature of human expertise is the ability to reason, a feature that induces a substantial amount of subjectivity. A viable expert system must model this phenomenon. Conventional systems use the theory of probability in their attempts to approximate human subjective reasoning.

The theory of probability has been the subject of controversy ever since its inception. The relation of probability to frequencies is often denied, and scholars have, therefore, broken down probability into two distinct parts: a part belonging to the realm of randomness which retains its relationship to frequencies, and a second part which is due to knowledge, and is known *a priori*. However, the additivity of probability has not been the subject of debate, and most discussions have tended to rely on an additive conception of probability.

The axiom of additivity states that the probability of a proposition and its opposite must sum to one. Stated simply, this means that for two independent events, the combined probability is exactly equal to the sum of the two individual probabilities. This constraint comes about as a result of the frequentative interpretation and should be reserved for the domain of pure chance. While probabilities defined by this axiom would suffice for the study of coin-tossing experiments, etc., it would seem

to be incorrect to extend this to the domain of evidence and its subjective combination. Non-additivity, on the other hand, implies that the combined probability of two independent propositions could be greater or less than (or even equal to) the sum of the individual proposition probabilities. Given the enigmatic nature of human judgments, it is reasonable to use non-additivity to mirror human strategies involved in the combination of evidence, and also utilities in economic theory.

In this chapter, we attempt to shed some light on the conundrum by tracing the historical conceptions of probability. We go on to examine monotonic measures (a general definition which includes additivity as well as non-additivity) in some detail, and finally, focus on a functional defined on monotonic measures.

HISTORICAL CONCEPTIONS OF PROBABILITY

The word *probability* is used today in a variety of ways, and philosophers have discerned many different *kinds* of probability. However, the most common, and perhaps, the most fundamental distinction is between *aleatory* (Latin, *aleae*: die, chance) and *epistemic* (Greek, *episteme*: knowledge) probabilities.

An aleatory (or, objective) probability of an outcome is simply the probability of a chance event and attempts to measure the propensity of its occurrence. Since this concept is approximated by the frequency with which the outcome does occur when a large number of trials are performed, it is a feature of the objective world. Due to this relationship with frequencies, aleatory probabilities must be additive. Epistemic probability, on the other hand, is strictly a feature of our knowledge. It is a number that very subjectively represents the degree to which we are certain of a proposition, the measure of our belief in it, or, the extent to which our evidence supports it. There is no necessary relation to frequencies, and therefore, epistemic probabilities need not be additive.

The foregoing definitions may appear to be idealized views of the overall conception of probability, and nuances in the way we understand probability could be used to attack them. But these nuances must not be allowed to obscure the important fact that at least a part of probability is a feature of knowledge and is due to nothing else. The view that epistemic probabilities should also be additive is deeply engrained in

current thought. This could be the result of a failure in recognizing the difference between aleatory and epistemic probabilities, and from a misunderstanding of the mathematics of additive probability. More importantly, the Bayesian theory of statistical inference has exerted considerable influence in favor of additivity. To gain insight into the debate on additivity, it is of interest to trace the historical conceptions of probability.

For several centuries, the idea of chance and the concept of belief have been united under the name probability. In this essay, we reject the unification and use the term probability to refer to the domain of subjective judgments and beliefs. Chance has been reserved for the realm of randomness.

According to Van Brakel (1976), the Greeks divided knowledge into three categories: "(i) that of which certain knowledge is possible, (ii) that of which probable knowledge is possible, and (iii) that of which no knowledge is possible." The first two categories arise out of Plato's distinction between knowledge (*episteme*: Latin, *scientia*) and opinion (*doxa*: Latin, *opinio*). Since the Greeks subscribed to determinism, the third category belonged to the realm of randomness. These distinctions appear to have endured through the middle ages; probability was an attribute of opinion where the random was quite out of play.

With the advent of the Renaissance, these categories were transformed for reasons still not clear. Hacking (1975) traces the transformation to the notion of *sign*, as it had been understood during

the middle ages. The modern concept of evidence was lacking, and the notion of *testimony* was extended by including signs - the *testimony of nature*. Opinion was based on testimony, and a probable opinion was one approved by some authority after observing the relevant signs.

Towards the end of the Renaissance, the connection between probability and chance seems to have first been made in a discussion of the philosophical concept of probability by Arnauld (1662). He distinguished between two kinds of evidence: *external evidence*, or the evidence of testimony, and *internal evidence*, the evidence of things. At that time, the mathematical theory of chance was just emerging, and Arnauld suggested that the principles of the new theory be used when considering the "probabilities of gain and loss in everyday life". Hacking (1975) is of the opinion that the origin of the new concept of internal evidence in the older concept of sign was reflected in a tendency of philosophers of the day to relate issues of evidence and probability to wagers in games of chance. This new kind of evidence made propositions worthy of approval by virtue of the frequency with which they made correct predictions. Although Arnauld used the theory of chance to calculate his probabilities, it is clear that they were epistemic. They were known *a priori* and, therefore, were unequivocally a feature of knowledge. The connection seems to have been introduced in an attempt to lend mathematical formality to the study of epistemic probability.

In the late seventeenth century, the tendency to associate epistemic ideas with chance could also have been furthered by the

activities of the practical statisticians of the age - the demographers. Stimulated by the new theory of games of chance, political curiosity, and the fashion of selling annuities, many authors began using a more epistemic vocabulary than had been seen in the early theory of chance. Yet, none of these authors followed Arnauld in using the weightier epistemic term - probability.

Jacob Bernoulli was the first substantial contributor to the theory of games of chance to grapple with its connection with probability. In his *Ars Conjectandi* (the Art of Conjecture), written around 1692, and published posthumously in 1713, Bernoulli has stated that probability is a degree of subjective certainty. The probabilities of different arguments have been combined to produce a probability based on the total evidence. Although Bernoulli has used methods from the mathematical theory of chance, non-additivity is evident, in several instances, when the probability of a thing and its opposite do not add to one. In essence, Bernoulli realized that combinations of arguments in the epistemic domain were quite different from corresponding manipulations in the theory of chance.

Bernoulli's subtle view of the connection between probability and chance did not endure, and Shafer (1978) has discerned important reasons for its failure. The theory of combining arguments was a prelimnary attempt and could not be compared as a mathematical theory with the already well-developed theory of games of chance. Bernoulli's understanding was a bit too subtle, and his successors simplified it by connecting his probability with the *ease of happening* as understood

in games of chance. The simplification was to some extent encouraged by Bernoulli's own *Law of Large Numbers*. This theorem which plays an important role in the theory of chance, maintains that in cases when the ease of happening of an event is not known *a priori*, it may be learned *a posteriori* from the observation of frequencies. Bernoulli thought one could use frequencies to find the ease of happening of various cases in individual arguments, the probabilities of these individual arguments could then be calculated and combined according to general rules. His successors abandoned his struggle with the combination of arguments and tended to think of probability as an ease of happening to be found directly from frequencies.

The word probability continued to have its broad epistemic connotations after Bernoulli's death. But the connection with chance gradually came to dominate the thinking of those who endeavored to treat epistemic probability numerically. In the works of Montmort (1708) and DeMoivre (1711), the notion of numerical probability essentially narrows to the paradigm of chance, and attempts to compute probabilities in situations other than games of chance are seen as extensions of the paradigm to those cases.

By the middle of the eighteenth century, the synthesis of probability and chance was complete. Lambert (1764) stands out as the only scholar at the time who was able to break away from the assimilation of probability in the additive theory of chance. He explicitly recognized and sought to explain the possible non-additivity of the probabilities of propositions, and he extended Bernoulli's rules

for combining arguments. Lambert's rule of combination turns out to be a special case of Dempster's rule of combination (Dempster, 1967).

Lambert's ideas did not influence the opinion at the time. Probability and chance continued to be used synonymously, and scholars began to learn, From Bayes (1764), Condorcet (1785), and Laplace (1785), just how additivity worked in the case of propositions.

Almost two centuries later, we accept the synthesis of probability and chance in the all-embracing term *probability*, and have further split it into aleatory and epistemic categories. More importantly, our study of epistemic probability has been pervaded by a universal and unconscious acceptance of additivity. The recent interest in AI, and its applications in thinking machines requires us to lessen our dependence on the restrictive constraint of additivity. For the sake of future progress, we may have to rediscover the concept of non-additivity introduced by Bernoulli and Lambert.

The reader is referred to Shafer (1978), Hacking (1975), Van Brakel (1976), and, Pearson and Kendall (1970) for interesting discussions on the history of probability.

MONOTONIC MEASURES

Kolmogorov (1933) was the first to axiomatize probability by representing random events as sets. According to this framework, probability is a normed measure defined on these sets. The measure-theoretic treatment has provided a logical and consistent foundation for the theory of probability and has united it with the mainstream of modern mathematics. In this section, the concept of a *monotonic measure*, defined using the same approach, is shown to include additive as well as non-additive features. We start by defining some basic concepts in measure theory.

Definition 3.2.1

A *σ-additive field* (or a *σ-algebra*) is a non-empty class of subsets of a set X which is closed under the formation of countable unions and complements and contains the empty set $\emptyset$.

Example 3.2.1

Let X be a finite set given by

$$X = \{x_1, x_2, x_3\}.$$

The smallest σ-field that can be generated is given by

$$\Sigma_o = \{X, \emptyset\}$$
$$= \{\{x_1, x_2, x_3\}, \emptyset\}.$$

The largest σ-field is

$$\Sigma_m = \{\{x_1, x_2, x_3\}, \emptyset, \{x_1, x_2\}, \{x_3\}, \{x_2, x_3\}, \{x_1\}, \{x_1, x_3\}, \{x_2\}\}.$$

Σ_m is also the power set of the finite set X, written as P(X).

Definition 3.2.2

A *Borel σ-additive field*, β, defined on any subset of X satisfies the following conditions:

i) $\beta \subseteq P(X)$ (3.1a)

ii) $\emptyset \in \beta$ (3.1b)

iii) If $Q \in \beta$, then $\overline{Q} \in \beta$ (3.1c)

iv) If $\forall i \in N$ (the set of natural numbers) $Q_i \in \beta$, then $\bigcup_{i \in N} Q_i \in \beta$. (3.1d)

Definition 3.2.3

The *Borel σ-field of subsets on the real line, R,* is the σ-field generated by the class of all bounded semi-closed intervals of the form (a, b], and is denoted by β_r.

Definition 3.2.4 (Sugeno, 1974)

A *monotonic* (or *fuzzy*) *measure*, g, is a fuction from a Borel field to [0,1], which has the following properties:

i) $g(\emptyset) = 0$, $g(X) = 1$; (Boundedness and Non-negativity) (3.2a)

ii) $\forall Q_1, Q_2 \in \beta$, if $Q_1 \subseteq Q_2$, then $g(Q_1) \leq g(Q_2)$; (Monotonicity) (3.2b)

iii) If $\forall i \in N$, $Q_i \in \beta$, and the sequence $(Q_i)_i$ is monotonic (i.e., $Q_1 \subseteq Q_2 \subseteq Q_3 \subseteq \ldots \subseteq Q_i \subseteq \ldots$, or, $Q_1 \supseteq Q_2 \supseteq Q_3 \supseteq \ldots \supseteq Q_i \supseteq \ldots$), then

$$\lim_{i\to\infty} g(Q_i) = g(\lim_{i\to\infty} Q_i); \text{ (Continuity).} \tag{3.2c}$$

Note that in these definitions, $\emptyset$ is the null or empty set, and X is the reference set.

Definition 3.2.5

(X , ß) is called a *measurable space*, or *Borel space*.

Definition 3.2.6

The triplet (X, ß, g) is known as a *monotonic* (or *fuzzy*) *measure space*.

The concept of a monotonic measure broadly defines most commonly used measures. The key axiom is monotonicity, property (ii) in Definition 3.2.4. This is a very general property that includes the lesser constraint of additivity. Hence, a probability (additive) measure is a member of the class of monotonic measures. The measure, g, is associated with a non-located element x_i of X. Sugeno (1974) has called g(Q) a *grade of fuzziness* of set Q. It expresses an evaluation of the statement

x_i belongs to Q

in a situation in which one subjectively guesses whether x_i is within Q (For the case of probability, the study of frequencies helps determine the grade which is objective and necessarily additive.). Thus, monotonicity of the measure g entails that

$$x_i \in Q_1$$

is never more certain than

$$x_i \in Q_2$$

when

$$Q_1 \subseteq Q_2.$$

It is also obvious that

$$\forall Q_1, Q_2 \in \beta, \; g(Q_1 \cup Q_2) \geq \mathrm{Max}(g(Q_1), g(Q_2)), \tag{3.3a}$$

and

$$\forall Q_1, Q_2 \in \beta, \; g(Q_1 \cap Q_2) \leq \mathrm{Min}(g(Q_1), g(Q_2)). \tag{3.3b}$$

For the case of a finite reference set, the continuity axiom, property (iii) in Definition 3.2.4, is dropped. It is also common to define the measure, g, on the power set, P(X).

We shall now examine some of the more commonly used monotonic measures defined over normal sets.

<u>Definition 3.2.7</u>

A monotonic measure, p, is a *probability measure* iff

i) $\forall\, i \in N$, $Q_i \in \beta$ and $\forall i \neq j$, $Q_i \cap Q_j = \emptyset$, then

$$p\Big(\bigcup_{i \in N} Q_i\Big) = \sum_{i \in N} p(Q_i); \text{ (Additivity).} \tag{3.4}$$

<u>Example 3.2.2</u>

Suppose we have an urn containing three balls of different colors, red, blue, and green. The balls collectively represent a set X, where

$X = \{x_1, x_2, x_3\}$,

$x_1 \equiv$ red ball,

$x_2 \equiv$ blue ball,

and

$x_3 \equiv$ green ball.

Let Q be any subset of X, i.e., $Q \subseteq X$, and assume that we randomly pick one colored ball from the urn.

a) If Q is the empty set, we know that the ball is not contained in set Q, and we write

$$p(\emptyset) = 0.$$

b) If Q is the reference set, X, the ball definitely belongs to the set, and,

$$p(X) = 1.$$

c) Suppose

$$Q_1 = \{x_1\}.$$

Since the ball has been selected at random, it may or may not be contained in set Q_1. Additionally, there is an equal chance that the ball could be either red, blue, or green. Hence, we assign a probability measure of 1/3 to set Q_1. In other words, we are 1/3 certain that the unknown ball belongs to Q_1. Similarly, we are also 1/3 certain that the ball belongs to each of

$$Q_2 = \{x_2\},$$

and

$$Q_3 = \{x_3\}.$$

Let

$$Q_4 = \{x_1, x_2\},$$

and we would like to assign a measure to which we are certain that

the unknown ball belongs to this set. Again, since each color is equally probable, the measure assigned to Q_4 is 2/3. Additivity is inherent in the choice of measures, and is due to the element of randomness. By Definition 3.2.7, we see that

$$
\begin{aligned}
p(Q_4) &= p(\{x_1, x_2\}) \\
&= p(\{x_1\} \cup \{x_2\}) \\
&= p(Q_1 \cup Q_2) \\
&= p(Q_1) + p(Q_2) \\
&= 1/3 + 1/3 \\
&= 2/3.
\end{aligned}
$$

Since we are certain that a ball is either contained in a set or its complement, we have

$$
\begin{aligned}
p(Q) + p(\overline{Q}) &= p(Q \cup \overline{Q}) \\
&= p(X) \\
&= 1.
\end{aligned}
$$

This example illustrates two important features of additivity, namely, the combined measure is exactly equal to the sum of the individual parts, and, the measure of a set added to its complement is exactly equal to one.

Definition 3.2.8 (Dubois and Prade, 1980)

A *Dirac measure* is a monotonic measure, d_i, defined by

$$Q \in \beta, \quad d_i(Q) = 1, \quad \text{iff } x_i \in Q \tag{3.5a}$$

$$= 0, \quad \text{otherwise} \tag{3.5b}$$

where x_i is a given element in X, $d_i(Q)$ is simply the membership

(or characteristic function) of x_i in a subset Q of X.

Example 3.2.3

Let

$$X = \{x_1, x_2, x_3\}.$$

a) If $Q = \emptyset$; $\forall x_i \in X$, $d_i(Q) = 0$.

b) If $Q = X$; $\forall x_i \in X$, $d_i(Q) = 1$.

c) Suppose that

$$Q_1 = \{x_1\},$$
$$Q_2 = \{x_2\},$$

and

$$Q_3 = \{x_3\}.$$

Then we have

$$d_1(Q_1) = 1, \quad d_2(Q_1) = 0, \quad d_3(Q_1) = 0,$$
$$d_1(Q_2) = 0, \quad d_2(Q_2) = 1, \quad d_3(Q_2) = 0,$$
$$d_1(Q_3) = 0, \quad d_2(Q_3) = 0, \quad d_3(Q_3) = 1.$$

It can be seen that d_i is additive for fixed i, e.g.,

$$\begin{aligned} d_1(\{x_1, x_2\}) &= d_1(\{x_1\} \cup \{x_2\}) \\ &= d_1(Q_1 \cup Q_2) \\ &= 1 \\ &= 1 + 0 \\ &= d_1(Q_1) + d_1(Q_2) \\ &= d_1(\{x_1\}) + d_1(\{x_2\}). \end{aligned}$$

Similarly,

$$d_2(\{x_1, x_2\}) = d_2(\{x_1\} \cup \{x_2\})$$

$$\begin{aligned}
&= d_2(Q_1 \cup Q_2)\\
&= 1\\
&= 0 + 1\\
&= d_2(Q_1) + d_2(Q_2)\\
&= d_2(\{x_1\}) + d_2(\{x_1\}),
\end{aligned}$$

and

$$\begin{aligned}
d_1(Q_1 \cup \overline{Q}_1) &= d_1(X)\\
&= 1\\
&= 1 + 0\\
&= d_1(Q_1) + d_1(\overline{Q}_1).
\end{aligned}$$

It is of interest to note that the Dirac assignment is performed when the color of the ball (Example 3.2.2) which has been picked from the urn is already known.

We now enter the realm of non-additivity which, in our opinion, is suitable for the treatment of epistemic probability.

Definition 3.2.9 (Zadeh, 1978)

A *possibility measure*, Π, is a monotonic measure such that for any collection $\{Q_i\}$ of subsets of X,

$$\Pi(\bigcup_i Q_i) = \operatorname*{Sup}_i \Pi(Q_i), \tag{3.6a}$$

and, for finite sets,

$$\Pi(\bigcup_i Q_i) = \operatorname*{Max}_i \Pi(Q_i). \tag{3.6b}$$

A possibility measure can be built from a *possibility distribution*,

i.e., a function π from X to [0,1] such that

$$\sup_{x_i \in X} \pi(x_i) = 1. \tag{3.7}$$

This is a normalization condition which implies that at least one event x_i is *absolutely possible*. The normalization condition also forces $\Pi(X)$ to be equal to one [property (i) in Definition 3.2.4].

Example 3.2.4

Consider the colored balls problem in Example 3.2.2. It is always *possible* that the unknown ball belongs to any non-empty subset of X, where

$$X = \{x_1, x_2, x_3\}.$$

Hence, we have

$$\pi(\{x_i\}) = 1, \quad \forall x_i \in X.$$

This leads to the assignment of the following possibility measures;

$$\Pi(Q) = 1; \quad \forall Q \subseteq X, \quad Q \neq \emptyset.$$

Let us suppose that for some (strange ?) reason, a person feels that the green ball plays hard to get. A visit to any casino will amply demonstrate this phenomenon. The assigned measures could be given by

$$\pi(\{x_1\}) = 1,$$

$$\pi(\{x_2\}) = 1,$$

and

$$\pi(\{x_3\}) = 0.5.$$

Thus

$$\Pi(Q) = 1; \quad \forall Q \subseteq X, \; Q \neq \emptyset, \; Q \neq \{x_3\},$$

$$\Pi(Q) = 0.5; \quad Q = \{x_3\},$$

and

$$\Pi(Q) = 0; \quad Q = \emptyset.$$

The possibility measure is not additive, and it is obvious that

$$\Pi(Q) + \Pi(\overline{Q}) > 1.$$

For further details concerning possibility measures, it is useful to examine Zadeh (1978), and Dubois and Prade (1980). Interesting relationships between possibility and *necessity* $[N(Q) = 1 - \Pi(\overline{Q})]$ measures are provided by Dubois and Prade (1983).

The measures discussed so far relate mainly to the concept of randomness (Possibility measures attempt to deal with subjectivity introduced in the analysis of randomness.). We shall now focus on measures that are useful for dealing with the subjectivity that is inherent in epistemic probability.

In a measure-theoretic treatment of epistemic probability, each element of the reference set, X, may be considered to be a fundamental proposition. A measure would then assign a number between zero and one to indicate a *degree of belief* or a grade of fuzziness accorded to a subset Q of X, on the basis of the evidence it contains. The reference set is given a measure of one, and this stands for the totality of evidence. Shafer (1976) has called it a *frame of discernment.*

The intuitive picture is as follows. A portion of belief committed to a proposition is also committed to any other proposition it implies. This means that a portion of belief accorded to a subset

is also accorded to any subset containing it. Thus, of the total belief committed to a subset Q of X, some may also be committed to one or more subsets of Q. However, there is a remainder that is committed exactly to Q - to Q and to no smaller subset. The fact that it ought to be possible to partition the total belief among different subsets of the frame of discernment, while assigning to each subset Q a portion that is committed to Q, and to nothing smaller, has led to the following definition.

Definition 3.2.10 (Shafer, 1976)

If X is a finite frame of discernment (reference set), then a function, m, from P(X) to [0,1] is called a *basic probability assignment* whenever

i) $m(\emptyset) = 0$ (3.8a)

and

ii) $\sum_{Q \subseteq X} m(Q) = 1.$ (3.8b)

Note that m(Q) measures the belief committed exactly to Q, not the total belief accorded to Q. Hence, to obtain the total belief committed to Q, we must add to m(Q), the quantities $m(Q_i)$ for all proper subsets Q_i of Q; i.e.,

$$Bel(Q) = \sum_{Q_i \subseteq Q} m(Q_i). \quad (3.9)$$

This is known as a *belief function* (Shafer, 1976). The following example should clarify the intuitive understanding of the belief function.

Example 3.2.5

Combining individual propositions into a body of evidence can be likened to building an edifice brick by brick. We use this idea to illustrate the intuitive aspects of a belief function.

Let us suppose that our edifice (the totality of evidence) consists of three bricks cemented together. This corresponds to the frame of discernment,

$$X = \{x_1, x_2, x_3\}.$$

In this model, the basic probability assignment, m, represents the masses of the individual elements of the structure. So we have,

$$m(\{x_1\}) \equiv \text{mass of brick 1},$$

$$m(\{x_2\}) \equiv \text{mass of brick 2},$$

and

$$m(\{x_3\}) \equiv \text{mass of brick 3}.$$

Suppose we start with brick 1. The total mass is given by

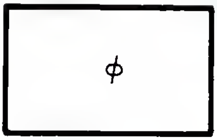
$$m(\emptyset) + m(\{x_1\}).$$

A wall does not consist of bricks alone. Cement is used to bind them together. Thus, when we join brick 2 to brick 1, the total mass is the sum of the masses of the individual bricks, plus an amount that is due to the cement bond. This is given by

$$m(\emptyset) + m(\{x_1\}) + m(\{x_2\}) + m(\{x_1, x_2\}).$$

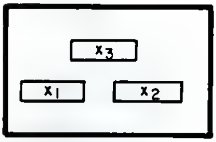
It is the mass of the cement bond alone, $m(\{x_1, x_2\})$, that corresponds to the portion of belief that is committed exactly to set Q, and to no smaller set, in Shafer's theory. This fits beautifully into our picture of evidence combination. Whenever we combine the weights of two separate

a)



In the beginning, there is emptiness.
$m(\emptyset) = 0;$
$Bel(\emptyset) = m(\emptyset)$
$= 0.$

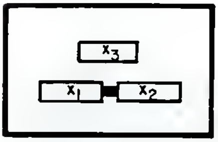
b)



Three bricks are introduced into the emptiness.
$m(\{x_i\}) > 0,\ i = 1, 2, 3;$

$Bel(\{x_i\}) = m(\emptyset) + m(\{x_i\}).$

c)

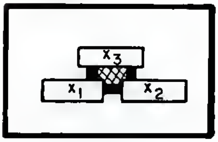


Two bricks are joined using a small amount of cement, $m(\{x_1, x_2\})$.

$$Bel(\{x_1, x_2\}) = m(\emptyset) + m(\{x_1\}) + m(\{x_2\}) + m(\{x_1, x_2\}).$$

$$Bel(\{x_3\}) = m(\emptyset) + m(\{x_3\}).$$

d)



To complete the edifice, it is necessary to join the third brick to the other two, using some more cement.

$$Bel(X) = m(\emptyset) + m(\{x_1\}) + m(\{x_2\}) + m(\{x_3\}) + m(\{x_1, x_2\}) + m(\{x_1, x_3\}) + m(\{x_2, x_3\}) + m(\{x_1, x_2, x_3\}).$$

$Bel(X) = 1$ (by definition).

Figure 3.1. Combination of evidence using the belief function.

propositions, we have an additional weight that is the result of the combination - the glue that is an essential part of the union. Figure 3.1 demonstrates the steps involved in building the edifice.

Definition 3.2.11 (Shafer, 1976)

A monotonic measure defined on a finite set, X, is a *belief function* iff

$$\forall Q_1, Q_2, \ldots, Q_n \in P(X),$$

$$\text{Bel}(\bigcup_i Q_i) \geq \sum_{i=1}^{n} \text{Bel}(Q_i) - \sum_{i<j} \text{Bel}(Q_i \cap Q_j) + \ldots$$

$$\ldots + (-1)^{n+1}\,\text{Bel}(\bigcap_i Q_i). \tag{3.10}$$

Note that the belief function is defined on the monotonic measure space, (X, P(X), Bel). The reference set, X, is always assumed to be finite in Shafer's theory of evidence (Shafer, 1976).

Example 3.2.6

Let

$$X = \{x_1, x_2, x_3\},$$

and

$$Q = \{x_1, x_2\}.$$

Due to the definition of the belief function in terms of the basic probability assignments [Equation (3.9)], we have

$$\begin{aligned}\text{Bel}(X) &= m(\emptyset) + m(\{x_1\}) + m(\{x_2\}) + m(\{x_3\}) + m(\{x_1, x_2\}) \\ &\quad + m(\{x_1, x_3\}) + m(\{x_2, x_3\}) + m(\{x_1, x_2, x_3\}) \\ &= 1,\end{aligned}$$

$$\mathrm{Bel}(Q) = m(\emptyset) + m(\{x_1\}) + m(\{x_2\}) + m(\{x_1, x_2\}),$$

and

$$\mathrm{Bel}(\overline{Q}) = \mathrm{Bel}(\{x_3\})$$

$$= m(\emptyset) + m(\{x_3\}).$$

Hence, if

$$m(Q_i) > 0; \ \forall Q_i \subseteq X, \ Q_i \neq \emptyset,$$

we have

$$\mathrm{Bel}(Q) + \mathrm{Bel}(\overline{Q}) < 1.$$

In general, however,

$$\mathrm{Bel}(Q) + \mathrm{Bel}(\overline{Q}) \leq 1.$$

This means that a lack of belief in an unlocated $x_i \in Q$, does not imply a strong belief in $x_i \in \overline{Q}$.

For further details concerning belief functions, see Shafer (1976).

Shafer's belief functions, while very general in their scope, do not lend themselves to the specification of functionals defined over them. This is due to the fact that knowledge of $\mathrm{Bel}(Q_1)$ and $\mathrm{Bel}(Q_2)$ is not always sufficient to calculate $\mathrm{Bel}(Q_1 \cup Q_2)$ [see Equation (3.10)].

In order to define a functional on the monotonic measure space, (X, β, g), the values of g must be given over the entire domain. Additionally, since g is monotonic, this property must be satisfied by all members of β without exception. Suppose that X is a finite set with n elements, and P(X) is taken to be β, the number of monotone sequences in β is n!. Thus, without a simple rule to define the measure, g, it is almost impossible to proceed.

Sugeno (1974) has provided a function that specifies g over the entire domain:

$$g(Q_1 \cup Q_2) = f(g(Q_1), g(Q_2)), \tag{3.11a}$$

where

$$Q_1, Q_2 \in \beta; Q_1 \cap Q_2 = \emptyset, \tag{3.11b}$$

and

$$f(y_1, y_2) = y_1 + y_2 + \lambda \cdot y_1 \cdot y_2 \; ; \lambda \in (-1, \infty). \tag{3.11c}$$

The definition of a Sugeno measure follows.

Definition 3.2.12 (Sugeno, 1974)

A monotonic measure, g_λ, is a *Sugeno measure* iff

$$\forall Q_1, Q_2 \in \beta; Q_1 \cap Q_2 = \emptyset,$$

$$g_\lambda(Q_1 \cup Q_2) = g_\lambda(Q_1) + g_\lambda(Q_2) + \lambda \cdot g_\lambda(Q_1) \cdot g_\lambda(Q_2), \tag{3.12a}$$

where

$$\lambda \in (-1, \infty). \tag{3.12b}$$

More generally, when Q_1 and Q_2 are any subsets of X, the following formula (Dubois and Prade, 1980) holds;

$$g_\lambda(Q_1 \cup Q_2) = \frac{g_\lambda(Q_1) + g_\lambda(Q_2) - g_\lambda(Q_1 \cap Q_2) + \lambda \cdot g_\lambda(Q_1) \cdot g_\lambda(Q_2)}{1 + \lambda \cdot g_\lambda(Q_1 \cap Q_2)}, \tag{3.13a}$$

where

$$\lambda \in (-1, \infty). \tag{3.13b}$$

Note that unlike the belief functions of Shafer, the Sugeno measure is not restricted to finite reference sets.

Example 3.2.7

It is interesting to see how the restriction, $\lambda \in (-1, \infty)$, in Equations (3.11c), (3.12b), and (3.13b), arises.

a) It is obvious that the following relation should always hold.

$$g_\lambda(\bar{\bar{Q}}) = g_\lambda(Q); \ \forall Q \subseteq X.$$

Now,

$$\begin{aligned} g_\lambda(Q \cup \bar{Q}) &= g_\lambda(X) \\ &= 1 \\ &= g_\lambda(Q) + g_\lambda(\bar{Q}) + \lambda \cdot g_\lambda(Q) \cdot g_\lambda(\bar{Q}). \end{aligned}$$

Hence,

$$g_\lambda(\bar{Q}) = \frac{1 - g_\lambda(Q)}{1 + \lambda \cdot g_\lambda(Q)},$$

and thus,

$$\begin{aligned} g_\lambda(\bar{\bar{Q}}) &= \frac{1 - g_\lambda(\bar{Q})}{1 + \lambda \cdot g_\lambda(\bar{Q})}, \\ &= \frac{1 - \dfrac{1 - g_\lambda(Q)}{1 + \lambda \cdot g_\lambda(Q)}}{1 + \lambda . \dfrac{1 - g_\lambda(Q)}{1 + \lambda \cdot g_\lambda(Q)}} \\ &= \frac{1 + \lambda \cdot g_\lambda(Q) - 1 + g_\lambda(Q)}{1 + \lambda \cdot g_\lambda(Q) + \lambda - \lambda \cdot g_\lambda(Q)}, \end{aligned}$$

or,

$$g_\lambda(\bar{\bar{Q}}) = \frac{g_\lambda(Q) \ (1 + \lambda)}{(1 + \lambda)}.$$

This implies that

$$\lambda \neq -1.$$

b) For all

$$Q_1 \cap Q_2 = \emptyset$$

we have by Definition 3.2.12,

$$g_\lambda(Q_1 \cup Q_2) = g_\lambda(Q_1) + g_\lambda(Q_2) + \lambda \cdot g_\lambda(Q_1) \cdot g_\lambda(Q_2)$$

$$= g_\lambda(Q_1) + g_\lambda(Q_2) \cdot (1 + \lambda \cdot g_\lambda(Q_1)).$$

Also,

$$g_\lambda(Q_1 \cup Q_2) \geqslant g_\lambda(Q_1).$$

Hence, we obtain

$$g_\lambda(Q_2) \cdot (1 + \lambda \cdot g_\lambda(Q_1)) \geqslant 0.$$

Now, since

$$g_\lambda(Q_1),\ g_\lambda(Q_2) \in [0,1],$$

we can write

$$1 + \lambda \cdot g_\lambda(Q_1) \geqslant 0,$$

indicating that

$$\lambda \geqslant -1.$$

Thus, since [see part (a)]

$$\lambda \neq -1,$$

we have

$$\lambda \in (-1, \infty).$$

<u>Example 3.2.8</u>

a) If $\lambda = 0$, the Sugeno measure has the additive structure of a probability measure.

For Q_1, $Q_2 \varepsilon \beta$, $Q_1 \cap Q_2 = \emptyset$,

$$g_\lambda(Q_1 \cup Q_2) = g_\lambda(Q_1) + g_\lambda(Q_2) + \lambda \cdot g_\lambda(Q_1) \cdot g_\lambda(Q_2),$$

and when $\lambda = 0$,

$$g_o(Q_1 \cup Q_2) = g_o(Q_1) + g_o(Q_2).$$

b) Since

$$g_\lambda(\overline{Q}) = \frac{1 - g_\lambda(Q)}{1 + \lambda \cdot g_\lambda(Q)},$$

we obtain

$$g_\lambda(Q) + g_\lambda(\overline{Q}) = g_\lambda(Q) + \frac{1 - g_\lambda(Q)}{1 + \lambda \cdot g_\lambda(Q)}$$

$$= \frac{g_\lambda(Q) + \lambda \cdot g_\lambda^2(Q) + 1 - g_\lambda(Q)}{1 + \lambda \cdot g_\lambda(Q)},$$

or

$$g_\lambda(Q) + g_\lambda(\overline{Q}) = \frac{1 + \lambda \cdot g_\lambda^2(Q)}{1 + \lambda \cdot g_\lambda(Q)}.$$

Since,

$$g_\lambda(Q) \varepsilon [0,1],$$

it is obvious that

$$g_\lambda^2(Q) \leqslant g_\lambda(Q),$$

and thus,

i) for $\lambda \varepsilon [0, \infty)$,

$$g_\lambda(Q) + g_\lambda(\overline{Q}) \leqslant 1,$$

which means that a lack of belief in a non-located $x_i \varepsilon Q$ does not imply a strong belief in $x_i \varepsilon \overline{Q}$. Hence, in this range, the Sugeno measure is a belief function of Shafer.

ii) for $\lambda \varepsilon (-1, 0]$,

we have

$$g_\lambda(Q) + g_\lambda(\overline{Q}) \geqslant 1,$$

or, a lack of belief in a proposition implies a very strong belief in its negation.

Example 3.2.8 demonstrates that the Sugeno measure with

$$\lambda \varepsilon [0, \infty) \tag{3.14}$$

is intuitively similar to Shafer's belief function (Example 3.2.5). However, for

$$\lambda \varepsilon (-1, 0), \tag{3.15a}$$

there is a certain amount of overlap whenever two independent propositions are combined, i.e., for

$$Q_1 \cap Q_2 = \emptyset \tag{3.15b}$$

we have

$$g_\lambda(Q_1 \cup Q_2) < g_\lambda(Q_1) + g_\lambda(Q_2). \tag{3.15c}$$

This seems to model the phenomenon of marginal utility in economic theory. The marginal utility, or extent of increase in satisfaction per unit of commodity, in general, decreases with each increase in the amount of commodity consumed. On the other hand, the total utility, which follows the axiom of monotonicity, always increases.

The *plausibility* of a subset Q of a finite set X has been defined by Shafer (1976) as

$$P\ell(Q) = 1 - \mathrm{Bel}(\overline{Q}). \tag{3.16}$$

<u>Definition 3.2.13</u> (Shafer, 1976)

A *plausibility measure*, $P\ell$, is a monotonic measure for which

$$\forall Q_1, Q_2, \ldots, Q_n \in P(X),$$

$$P\ell(\bigcap_i Q_i) < \sum_{i=1}^{n} P\ell(Q_i) - \sum_{i<j} P\ell(Q_i \cup Q_j) + \cdots$$
$$\cdots + (-1)^{n+1} P\ell(\bigcup_i Q_i). \tag{3.17}$$

<u>Example 3.2.9</u>

a) Suppose that

$$Q_1, Q_2 \subseteq X, \quad Q_1 \cap Q_2 = \emptyset.$$

Then, by Definition 3.2.13,

$$P\ell(Q_1 \cap Q_2) \leqslant P\ell(Q_1) + P\ell(Q_2) - P\ell(Q_1 \cup Q_2),$$

or

$$P\ell(Q_1 \cup Q_2) \leqslant P\ell(Q_1) + P\ell(Q_2) - P\ell(Q_1 \cap Q_2)$$
$$\leqslant P\ell(Q_1) + P\ell(Q_2) - P\ell(\emptyset).$$

Thus,

$$P\ell(Q_1 \cup Q_2) \leqslant P\ell(Q_1) + P\ell(Q_2).$$

b) For any $Q \subseteq X$,

we have, using Equation (3.16),

$$P\ell(Q) + P\ell(\overline{Q}) = (1 - Bel(\overline{Q})) + (1 - Bel(Q))$$
$$= 2 - (Bel(Q) + Bel(\overline{Q})).$$

Since (see Example 3.2.6),

$$Bel(Q) + Bel(\overline{Q}) \leqslant 1,$$

we obtain

$$P\ell(Q) + P\ell(\overline{Q}) \geqslant 1.$$

Thus, a plausibility measure has the same structure as a Sugeno measure

with $\lambda \varepsilon (-1, 0]$.

The following theorems are pertinent.

Theorem 3.2.1

A Sugeno measure is a belief function iff

$$\lambda \varepsilon [0, \infty). \tag{3.18}$$

Proof (Banon, 1978)

Let Q be a subset of X finite. Developing $g_\lambda(Q)$ in terms of $g(\{x_i\})$'s yields,

$$g_\lambda(Q) = \sum_{Q_j \subseteq Q} \lambda^{\mathrm{Card}(Q_j) - 1} \prod_{x_i \varepsilon Q_j} g(\{x_i\}), \tag{3.19}$$

where $\mathrm{Card}(Q_j)$ is the cardinality of subset Q_j.

Thus, by writing

$$m(Q_j) = \lambda^{\mathrm{Card}(Q_j) - 1} \prod_{x_i \varepsilon Q_j} g(\{x_i\}); \quad \text{iff } \lambda > 0, \tag{3.20}$$

we see that

$$\begin{aligned} g_\lambda(Q) &= \sum_{Q_j \subseteq Q} m(Q_j) \\ &= \mathrm{Bel}(Q). \end{aligned} \tag{3.21}$$

Q.E.D.

Theorem 3.2.2

A Sugeno measure is a plausibility measure iff

$$\lambda \varepsilon (-1, 0]. \tag{3.22}$$

Proof (Dubois and Prade, 1980)

Let g be a Sugeno measure with $\lambda \varepsilon (-1, 0]$, and denote

$$f(Q) = 1 - g_\lambda(\overline{Q}). \tag{3.23}$$

For any $Q_1, Q_2 \subseteq X$,

$$f(Q_1 \cup Q_2) = 1 - g_\lambda(\overline{Q_1 \cup Q_2}). \tag{3.24}$$

On expressing $g_\lambda(\overline{Q_1 \cup Q_2})$ in terms of $g_\lambda(\overline{Q_1})$, $g_\lambda(\overline{Q_2})$, and $g_\lambda(\overline{Q_1 \cap Q_2})$, and simplifying, we obtain

$$f(Q_1 \cup Q_2) = \frac{f(Q_1) + f(Q_2) + \overline{\lambda}\cdot f(Q_1)\cdot f(Q_2) - f(Q_1 \cap Q_2)}{1 + \overline{\lambda}\cdot f(Q_1 \cap Q_2)}, \tag{3.25}$$

where

$$\overline{\lambda} = \frac{-\lambda}{1 + \lambda}. \tag{3.26}$$

Thus, f is a Sugeno measure with parameter $\overline{\lambda}$.

Note that the function

$$\lambda \rightarrow \frac{-\lambda}{1 + \lambda} \quad (= \overline{\lambda}) \tag{3.27}$$

is an involutive bijection from $(-1, 0]$ to $[0, \infty)$.

Due to the definition of plausibility measures in terms of belief functions, and also to the fact that g is a belief function iff $\lambda \in [0, \infty)$ [Theorem 3.2.1], a Sugeno measure is a plausibility measure iff $\lambda \in (-1,0]$.

Q.E.D.

AN ALGORITHM FOR THE IDENTIFICATION OF SUGENO MEASURES

In order to be able to use monotonic measures in applications, it is essential that they be specified over the entire domain. The Sugeno measure, due to its explicit definition, is easily fitted from researched data. This section describes an efficient algorithm (Wierzchon, 1983) for the determination of Sugeno measures by *regression* from subjective estimates obtained by experiment.

Let

$$X = \{x_1, x_2, \ldots, x_n\}, \tag{3.28}$$

be a finite reference set, and let

$$w(Q),\ Q \subseteq X, \tag{3.29}$$

be the experimentally determined subjective estimates which are to be fitted to the constraints of Sugeno measures. The procedure must look for densities

$$g_i = g(\{x_i\}),\ i = 1, 2, \ldots, n \tag{3.30}$$

with a corresponding value of λ so that an appropriately defined error function is minimized. In mathematical terms, it is customary to state the problem as:

Locate densities,

$$g_i,\ i = 1, 2, \ldots, n \tag{3.31a}$$

so as to minimize

$$J = \sum_{Q \subseteq X} (w(Q) - g_\lambda(Q))^2 \tag{3.31b}$$

subject to the Sugeno constraints on g_λ, i.e.,

$$g_\lambda(\emptyset) = 0, \tag{3.31c}$$

$$g_\lambda(X) = 1, \tag{3.31d}$$

and

$$\forall Q_1,\ Q_2 \in P(X),\ Q_1 \cap Q_2 = \emptyset,$$

$$g_\lambda(Q_1 \cup Q_2) = g_\lambda(Q_1) + g_\lambda(Q_2) + \lambda \cdot g_\lambda(Q_1) \cdot g_\lambda(Q_2). \tag{3.31e}$$

The optimization problem defined by Equation (3.31) can be solved using a suitable objective function minimization technique. Sekita and Tabata (1977) have suggested the use of a (rather tedious) Sequential Unconstrained Minimization Technique (SUMT). Other objective function minimization techniques can also be used. However, the method of Wierzchon (1983), presented in this section, is simple, fast and easily programmable. The method is based on the following important results.

<u>Theorem 3.3.1</u>

Let (X, β, p) be a measurable space with a probability or finite Lebesgue measure, p. A composition $f \circ p$ produces a Sugeno measure if f is of the form

$$f(y) = \frac{1}{\lambda}\ (\ c^y - 1\ );\quad c > 0,\ c \neq 1. \tag{3.32}$$

<u>Proof</u> [see Wierzchon, 1983.]

<u>Corollary 3.3.1</u>

A Sugeno measure g_λ defined on β produces exactly one probability measure p defined on this β where

$$p(Q) = \log_{(1+\lambda)}(1 + \lambda \cdot g_{\lambda}(Q)), \; Q \in \beta. \tag{3.33}$$

The inverse is not true [see Wierzchon, 1983].

Example 3.3.1

Let $p(Q_1)$ and $p(Q_2)$ be probability measures for

$Q_1, Q_2 \subseteq X$ (finite),

and

$Q_1 \cap Q_2 = \emptyset$.

a) By definition of probability measures (Definition 3.2.7) we have

$$p(Q_1 \cup Q_2) = p(Q_1) + p(Q_2).$$

Using Theorem 3.3.1 (subject to the restrictions on c), we obtain

$$f \bullet p(Q_1) = \frac{1}{\lambda}\left(c^{p(Q_1)} - 1 \right)$$

$$f \bullet p(Q_2) = \frac{1}{\lambda}\left(c^{p(Q_2)} - 1 \right)$$

and

$$f \bullet p(Q_1 \cup Q_2) = \frac{1}{\lambda}\left(c^{p(Q_1 \cup Q_2)} - 1 \right).$$

As shown below, f p is a Sugeno measure;

$$\begin{aligned} f \bullet p(Q_1 \cup Q_2) &= f \bullet p(Q_1) + f \bullet p(Q_2) + \lambda \cdot f \bullet p(Q_1) \cdot f \bullet p(Q_2) \\ &= \frac{1}{\lambda}(c^{p(Q_1)} - 1) + \frac{1}{\lambda}(c^{p(Q_2)} - 1) + \frac{1}{\lambda}(c^{p(Q_1)} - 1)(c^{p(Q_2)} - 1) \\ &= \frac{1}{\lambda}(c^{p(Q_1)} - 1 + c^{p(Q_2)} - 1 + c^{p(Q_1)+p(Q_2)} - c^{p(Q_1)} - c^{p(Q_2)} \\ &\quad + 1) \end{aligned}$$

$$= \frac{1}{\lambda} (c^{p(Q_1)+p(Q_2)} - 1)$$

$$= \frac{1}{\lambda} (c^{p(Q_1 \cup Q_2)} - 1)$$

$$= f \circ p(Q_1 \cup Q_2).$$

<u>Example 3.3.2</u>

It is of interest to see how the restrictions on c [Equation (3.32)] arise. From Theorem 3.3.1, we have

$$g_\lambda(Q) = \frac{1}{\lambda} (c^{p(Q)} - 1); \quad c > 0, \ c \neq 1, \ Q \subseteq X.$$

But, by definition of monotonic measures,

$$g_\lambda(X) = 1,$$

and

$$p(X) = 1.$$

On writing

$$g_\lambda(X) = \frac{1}{\lambda} (c^{p(X)} - 1),$$

and simplifying, we obtain

$$1 = \frac{1}{\lambda} (c^1 - 1).$$

Thus,

$$c = \lambda + 1.$$

However, by definition of Sugeno measures (Definition 3.2.12),

$$\lambda \in (-1, \infty],$$

and, therefore,

$$c > 0.$$

Also, since $\lambda = 0$ corresponds to the additive case,

$$c \neq 1.$$

Hence,

$$c > 0, \quad c \neq 1.$$

Since the minimization of the error function,

$$J = \frac{1}{m} \sum_{j=1}^{m} \left(w(Q_j) - g_\lambda(Q_j) \right)^2 \tag{3.34a}$$

where

$$Q_j \in P(X) \tag{3.34b}$$

and

$$m = \mathrm{Card}(P(X)), \tag{3.34c}$$

subject to the Sugeno constraints on g_λ, is rather difficult from a mathematical standpoint, the problem is simplified by employing the transformation provided by Corollary 3.3.1 to define a new error function;

$$R = \frac{1}{m} \sum_{j=1}^{m} \left(v(Q_j) - p(Q_j) \right)^2 \tag{3.35a}$$

where

$$v(Q_j) = \log_c(1 + \lambda \cdot w(Q_j)) \tag{3.35b}$$

$$\lambda = c - 1 \tag{3.35c}$$

$$Q_j \in P(X) \tag{3.35d}$$

and

$$m = \mathrm{Card}(P(X)). \tag{3.35e}$$

Wierzchon (1983) has also shown that minimizing the error function R [Equation (3.35a)] is equivalent to minimizing the function J [Equation (3.34a)], and the optimization problem simplifies to:

Locate probability densities,

$$p_i, \ i = 1, 2, \ldots, n \tag{3.36a}$$

so as to minimize

$$R = \frac{1}{m} \sum_{j=1}^{m} \left(v(Q_j) - p(Q_j) \right)^2 \tag{3.36b}$$

subject to the probability constraints,

$$p_i \geqslant 0, \ i = 1, 2, \ldots, n \tag{3.36c}$$

$$\sum_{i=1}^{n} p_i = 1, \tag{3.36d}$$

$$p(\emptyset) = 0, \tag{3.36e}$$

and

$$p(X) = 1. \tag{3.36f}$$

The method of least squares is employed to derive an analytic solution to the minimization problem.

The overall approach is as follows;

i) Reduce the identification procedure to the probability domain using the fact that the Sugeno measure is the exponential transformation of a probability measure (Corollarv 3.3.1).

ii) Minimize the error function using the method of least squares.

iii) Having identified the probability measures, transform them to Sugeno measures (Theorem 3.3.1).

The algorithm is comprised of the following steps [for details and derivations, see Wierzchon (1983)];

i) Read in the data: the cardinality n of the set, X; and experimentally obtained subjective weights, $w(Q_j)$, $j = 1, 2, \ldots, 2^{n-2}$, i.e.,

$Q_j \subseteq X$, $Q_j \neq \emptyset$, $Q_j \neq X$.

ii) Find the values of λ by solving the following equation;

$$\prod_{j=1}^{m} (1 + \lambda \cdot w(Q_j))^{k_j} = (\lambda + 1)^{2^{n-2}(n+1) - n} \qquad (3.37a)$$

where

$$k_j = \mathrm{Card}(Q_j) \qquad (3.37b)$$

$$m = \mathrm{Card}(P(X)) \qquad (3.37c)$$

$$n = \mathrm{Card}(X). \qquad (3.37d)$$

iii) Using a non-complex value of λ closest to zero, compute the values of z_i, $i = 1, 2, \ldots, n$, according to

$$z_i = \log_c \prod_{j=1}^{m} (1 + \lambda \cdot w(Q_j) \cdot d_{ij}), \qquad (3.38a)$$

where

$$d_{ij} = 1, \text{ if } x_i \in Q_j \qquad (3.38b)$$
$$= 0, \text{ otherwise.} \qquad (3.38c)$$

iv) Compute the values of p_i, $i = 1, 2, . . . , n$, according to the equation

$$p_i = 2^{2-n}(z_i + 1) - 1. \tag{3.39}$$

v) Finally, compute the Sugeno densities, g_i, $i = 1, 2, . . , n$, using the transformation,

$$g_i = \frac{1}{\lambda}(c^{p_i} - 1) \tag{3.40a}$$

where

$$c = \lambda + 1. \tag{3.40b}$$

The preceding algorithm has been termed a *direct method* for the estimation of Sugeno densities because it requires the knowledge of weights for all subsets of set X. There is, however, a practical difficulty. When the cardinality of the reference set is large, a considerable number of subjective estimates are needed (If there are 10 elements in the reference set, a subject has to provide 1022 estimates!). In such instances, we may drop from the set X those aspects whose grades of importance are close to zero. Lowering the cardinality results in a reduction of the number of subjective judgments that are required (If the number of elements in the reference set is reduced from 10 to 8, only 254 subjective estimates are needed.).

THE SUGENO INTEGRAL

In the preceding sections, the mathematical and intuitive aspects of monotonic measures have been reviewed in some detail. We have seen that measures with non-additivity have properties which could render them useful for modeling human decision making strategies. Before we venture any further, let us sum up the ideas we have expounded to this point.

There are two distinct types of probability, aleatory and epistemic. The measure-theoretic approach for aleatory probability represents random events as sets. A weight (or, probability) is assigned to each subset depending on the likelihood that an unknown event would belong to it. In a sense, the measure of probability could be taken to represent the *grade of importance* the specific subset has for the purpose of predicting an event. Due to an implicit relation between randomness and frequencies, the measure is necessarily additive.

In the epistemic domain, each element of a set is a fundamental proposition. A measure defined on a subset is taken to indicate the degree of belief accorded to the subset on the basis of evidence it contains. Observation of human behavior points to many non-compensatory and conjunctive strategies. It is these features that we attempt to model using non-additivity. The following example illustrates the point.

Example 3.4.1

For some time, a Chemical Engineering Department has been searching

for a suitable candidate to fill a vacant position of Assistant Professor in the department. The advertisement states that a young candidate with demonstrated research skills, and substantial teaching ability is preferred. Since the department is oriented towards research, the candidate's research skill is very important. However, it is not necessary that the candidate be young.

A mathematical formulation could represent the reference set by

$$X = \{x_1, x_2, x_3\}$$

where

$x_1 \equiv$ candidate is young,

$x_2 \equiv$ candidate is a good teacher,

and

$x_3 \equiv$ candidate is a good researcher.

Members of the selection committee were asked for their opinions concerning the relative weights (measures) for each subset of X. The consensus values provided were as follows;

$$g(\emptyset) = 0$$

$$g(\{x_1\}) = 0.10$$

$$g(\{x_2\}) = 0.20$$

$$g(\{x_3\}) = 0.50$$

$$g(\{x_1, x_2\}) = 0.55$$

$$g(\{x_1, x_3\}) = 0.75$$

$$g(\{x_2, x_3\}) = 0.95$$

$$g(\{x_1, x_2, x_3\}) = 1.$$

Observe that these measures satisfy the axiom of monotonicity, and are non-additive. Each value expresses the *level of satisfaction* that would arise if a candidate has the qualities contained in the corresponding subset of X. A value of one entails sure selection of the candidate, while a value of 0.5 could be taken to suggest that the candidate may or may not be selected. Note that good research skills alone will not ensure selection. However, a person with demonstrated research and teaching abilities is almost a cinch for the position.

In this example, the propositions are all assumed to be answered with either a *yes* or a *no*. But, in real-life situations, a candidate might satisfy each proposition partially - somewhere between a sure yes and an emphatic no. Monotonic measures alone will not suffice. As we shall see, the Sugeno Integral, a functional defined on a Sugeno measure space (or other monotonic measure spaces), could help solve this difficulty.

Definition 3.4.1 (Sugeno, 1974)

Let (X, β, g) be a monotonic measure space, and let $h:X \rightarrow [0,1]$ be a measurable function defined on X. A *Sugeno Integral* (or *Fuzzy Integral*) over $Q \in \beta$, of $h(x)$ with respect to a monotonic measure g is defined by

$$\oint_Q h(x) \circ g(.) = \sup_{\alpha \in [0,1]} (\mathrm{Min}(\alpha, g(Q \cap F_\alpha))), \tag{3.41a}$$

where

$$F_\alpha = \{x | h(x) > \alpha\}. \tag{3.41b}$$

In this work, we are concerned with finite reference sets, and it is not necessary to assume continuity of monotonic measures. The definition for this case follows.

Definition 3.4.2 (Sugeno, 1974)

Let (X, P(X), g) be a measurable space, and let $h: X \rightarrow [0,1]$ be a function defined on X. The *Sugeno Integral* over any set $Q \in P(X)$ is given by

$$\int_Q h(x) \bullet g(.) = \underset{F \in P(X)}{\mathrm{Max}} \left(\underset{x_i \in F}{\mathrm{Min}} (h(x_i)) \wedge g(Q \cap F) \right). \tag{3.42}$$

For further details, see Sugeno (1974).

The following example demonstrates the evaluation of the Sugeno Integral.

Example 3.4.2

Let X be a finite reference set given by

$$X = \{x_1, x_2, x_3\}.$$

The monotonic measures on (X, P(X), g) are

$g(\emptyset) = 0$

$g(\{x_1\}) = 0.2$

$g(\{x_2\}) = 0.3$

$g(\{x_3\}) = 0.4$

$g(\{x_1, x_2\}) = 0.6$

$g(\{x_1, x_3\}) = 0.7$

$g(\{x_2, x_3\}) = 0.9$

and

$$g(\{x_1, x_2, x_3\}) = 1.$$

Let the function h(x) be given by

$$h(x_1) = 0.5$$

$$h(x_2) = 0.6$$

$$h(x_3) = 0.8.$$

a) Evaluation of the Sugeno Integral over $Q = \{x_1, x_2\}$.

$$\oint_Q h(x)\bullet g(.) = \underset{F\varepsilon P(X)}{\text{Max}} \left(\underset{x_i \varepsilon F}{\text{Min}} (h(x_i)) \wedge g(Q \cap F) \right).$$

$$= \text{Max} \begin{bmatrix} \text{Min}(h(\emptyset)) \wedge g(Q \cap \emptyset), \quad F = \emptyset \\ \text{Min}(h(x_1)) \wedge g(Q \cap \{x_1\}), \quad F = \{x_1\} \\ \text{Min}(h(x_2)) \wedge g(Q \cap \{x_2\}), \quad F = \{x_2\} \\ \text{Min}(h(x_3)) \wedge g(Q \cap \{x_3\}), \quad F = \{x_3\} \\ \text{Min}(h(x_1), h(x_2)) \wedge g(Q \cap \{x_1, x_2\}), \quad F = \{x_1, x_2\} \\ \text{Min}(h(x_1), h(x_3)) \wedge g(Q \cap \{x_1, x_3\}), \quad F = \{x_1, x_3\} \\ \text{Min}(h(x_2), h(x_3)) \wedge g(Q \cap \{x_2, x_3\}), \quad F = \{x_2, x_3\} \\ \text{Min}(h(x_1), h(x_2), h(x_3)) \wedge g(Q \cap \{x_1, x_2, x_3\}), \quad F = X \end{bmatrix}$$

$$= \text{Max} \begin{bmatrix} 0.0 \wedge 0.0\ [g(\emptyset)] & = 0.0 \\ 0.5 \wedge 0.2\ [g(\{x_1\})] & = 0.2 \\ 0.6 \wedge 0.3\ [g(\{x_2\})] & = 0.3 \\ 0.8 \wedge 0.0\ [g(\emptyset)] & = 0.0 \\ 0.5 \wedge 0.6\ [g(\{x_1, x_2\})] & = 0.5 \\ 0.5 \wedge 0.2\ [g(\{x_1\})] & = 0.2 \\ 0.6 \wedge 0.3\ [g(\{x_2\})] & = 0.3 \\ 0.5 \wedge 0.6\ [g(\{x_1, x_2\})] & = 0.5 \end{bmatrix}$$

Thus,

$$\oint_Q h(x)\circ g(.) = \mathrm{Max}(0.0,\ 0.2,\ 0.3,\ 0.0,\ 0.5,\ 0.2,\ 0.3,\ 0.5)$$

$$= 0.5.$$

b) Evaluation of the integral over $X = \{x_1, x_2, x_3\}$.

$$\oint_X h(x)\bullet g(.) = \underset{F\varepsilon P(X)}{\mathrm{Max}} \left(\underset{x_i \varepsilon F}{\mathrm{Min}} (h(x_i)) \wedge g(X \cap F) \right)$$

$$= \underset{F\varepsilon P(X)}{\mathrm{Max}} \left(\underset{x_i \varepsilon F}{\mathrm{Min}} (h(x_i)) \wedge g(F) \right).$$

On substitution, we obtain

$$\oint_X h(x)\bullet g(.) = \mathrm{Max}(0.0,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.5,\ 0.6,\ 0.5)$$

$$= 0.6.$$

Note that, in general, we have

$$\oint_Q h(x)\circ g(.) \leqslant \oint_X h(x)\bullet g(.), \quad Q \subseteq X. \tag{3.43}$$

Defined on a monotonic measure space, the Sugeno Integral is is interesting functional that is analogous to the Lebesgue Integral (a well-known functional defined on additive measures). Sugeno (1974) has shown that if p is a probability measure (additive monotonic measure), defined on a reference set, X, then

$$\left| \oint_X h(x)\bullet p(.) - \int_X h(x)dp \right| \leqslant \frac{1}{4} \tag{3.44a}$$

where

$$\underset{X}{\fint} h(x) \circ p(.) \tag{3.44b}$$

is the Sugeno Integral evaluated over X, and

$$\int_X h(x)\,dp \tag{3.44c}$$

is the Lebesgue Integral (or probability expectation value). The relation presented in Equation (3.44) demonstrates that the Sugeno evaluation is at most 1/4 away from the probabilistic expectation. Hence, we could interpret the Sugeno Integral as representing a subjective expectation value in applications where the subjectivity has been grasped by monotonic measures. Some of the more important properties of Sugeno Integrals are given below.

i) $$0 \leq \fint h \circ g(.) \leq 1. \tag{3.45}$$

ii) $$\fint (a \vee h) \circ g(.) = a \vee \fint h \circ g(.), \quad a \in [0,1] \tag{3.46}$$

iii) $$\fint (a \wedge h) \circ g(.) = a \wedge \fint h \circ g(.), \quad a \in [0,1] \tag{3.47}$$

iv) $$\fint (h_1 \vee h_2) \circ g(.) \geq \fint h_1 \circ g(.) \vee \fint h_2 \circ g(.) \tag{3.48}$$

v) $$\fint (h_1 \wedge h_2) \circ g(.) \leq \fint h_1 \circ g(.) \wedge \fint h_2 \circ g(.) \tag{3.49}$$

vi) $$\underset{Q_1 \cup Q_2}{\fint} h \circ g(.) \geq \underset{Q_1}{\fint} h \circ g(.) \vee \underset{Q_2}{\fint} h \circ g(.) \tag{3.50}$$

vii) $$\underset{Q_1 \cap Q_2}{\fint} h \circ g(.) \leq \underset{Q_1}{\fint} h \circ g(.) \wedge \underset{Q_2}{\fint} h \circ g(.) \tag{3.51}$$

These properties follow from the definition of the Sugeno Integral (Definitions 3.4.1 and 3.4.2), and are easy to prove. For further

details, see Sugeno (1974), and, Terano and Sugeno (1975).

The Sugeno Integral in Definitions 3.4.1 and 3.4.2 is just one functional defined on monotonic measures. It is possible to define different functionals that may be suitable for other applications. We are concerned with modeling human subjectivity, and from this standpoint, the most important property of Sugeno Integrals is monotonicity.

If

$$\forall x_i \in X,\ h_1(x_i) \leq h_2(x_i), \tag{3.52a}$$

then

$$\oint_X h_1(x) \circ g(.) \leq \oint_X h_2(x) \circ g(.), \quad \text{(monotonicity for the integrand, h)} \tag{3.52b}$$

Additionally,

$$\forall Q_1,\ Q_2 \in P(X),\ \ Q_1 \subseteq Q_2, \tag{3.53a}$$

$$\oint_{Q_1} h(x) \circ g(.) \leq \oint_{Q_2} h(x) \circ g(.), \quad \text{(monotonicity for the sets over which integration is performed)} \tag{3.53b}$$

These properties follow from the definition of the Sugeno Integral. We shall shortly see that monotonicity is essential for approximating human evaluative tendencies.

As stated previously, in the present framework, each element of the reference set, X, is a fundamental proposition or criterion. Any subset Q of this set consists of a collection of criteria, or intuitively, is an aspect or view of the overall picture. A monotonic measure provides a grade of importance, g(Q), to each view, Q. In general, a manifestation

would not satisfy the propositions completely. So the integrand, $h(x_i)$, represents a *truth value* in the sense of logic, or a level of satisfication of the proposition, x_i, with reference to the manifestation. The operation

$$\underset{x_i \varepsilon Q}{\text{Min}} \; (h(x_i)), \qquad (3.54)$$

would then provide the pessimistic or most secure level of satisfaction that the manifestation offers when examined from the point of view of the criteria contained in view Q.

The Sugeno Integral attempts to combine the most secure level of satisfaction obtained from a view, with the relative importance of that particular view. This is done for each view, Q, of the overall picture, X. The power set, P(X), lists all possible views. Finally, the integral gives the *mean* or expected value after considering all views.

There are two specific operations that are performed. First, the most secure level of satisfaction is combined with the relative importance of the corresponding view using the minimum operator. Since the value obtained can never be greater than the importance of the view, g(Q), this operation serves in limiting the evaluation offered by the view to a value no greater than its importance, g(Q).

The second operation selects the best evaluation from among all possible views. This is a common tendency in human judgments. We evaluate things from many different angles or aspects. The angle that *strikes* in terms of satisfaction as well as importance, plays

a major role in our final analyses.

The Sugeno Integral restricts itself to the use of "Max" and "Min" operators. Since both the level of satisfaction and the grade of importance take values between zero and one, these operators are appropriate, and the functional takes a non-linear form. Of course, we could also define isomorphic functionals that employ other operators. For example, the Lebesgue Integral uses "+" and "x" operators. These operators are deemed necessary to deal with additivity and the implicit relation between aleatory probability and frequencies.

The following example should clarify the intuitive aspects of the Sugeno Integral.

Example 3.4.3

As seen in Example 3.4.1, monotonic measures alone are not sufficient for selecting a candidate for the post of Assistant Professor. In this example, the Sugeno Integral is employed to arrive at a more meaningful decision.

The criteria are represented by

$$X = \{x_1, x_2, x_3\}$$

where

$x_1 \equiv$ candidate is young

$x_2 \equiv$ candidate is a good teacher

and

$x_3 \equiv$ candidate has good research abilities.

The relative weights for each collection of criteria, as given by the selection panel, are as follows;

$$g(\emptyset) = 0$$

$$g(\{x_1\}) = 0.10$$

$$g(\{x_2\}) = 0.20$$

$$g(\{x_3\}) = 0.50$$

$$g(\{x_1, x_2\}) = 0.55$$

$$g(\{x_1, x_3\}) = 0.75$$

$$g(\{x_2, x_3\}) = 0.95$$

and

$$g(\{x_1, x_2, x_3\}) = 1.$$

The weights satisfy the axiom of monotonicity; they indicate that a candidate with good research abilities is preferred, while it is not very important that the candidate must be young.

Let us suppose that two candidates have applied for the position, and the selection committee has felt that they have satisfied the criteria to the following levels.

$$h_1(x_1) = 0.4$$

$$h_1(x_2) = 0.7$$

and

$$h_1(x_3) = 0.8,$$

for candidate 1, and,

$$h_2(x_1) = 1.0$$

$$h_2(x_2) = 0.8$$

and

$h_2(x_3) \;=\; 0.4,$

for candidate 2.

a) Evaluation of candidate 1.

$$\oint_X h_1(x) \circ g(.) \;=\; \text{Max} \begin{bmatrix} 0.00 & \wedge & 0.00 \\ 0.40 & \wedge & 0.10 \\ 0.70 & \wedge & 0.20 \\ 0.80 & \wedge & 0.50 \\ 0.40 & \wedge & 0.55 \\ 0.40 & \wedge & 0.75 \\ 0.70 & \wedge & 0.95 \\ 0.40 & \wedge & 1.00 \end{bmatrix}$$

and we obtain

$$\oint_X h_1(x) \circ g(.) \;=\; 0.70.$$

Note that this value is obtained because the candidate has satisfied the important important criteria of research ability and teaching skill collectively to a value not lower than 0.7.

b) Evaluation of candidate 2.

$$\oint_X h_2(x) \circ g(.) \;=\; \text{Max} \begin{bmatrix} 0.00 & \wedge & 0.00 \\ 1.00 & \wedge & 0.10 \\ 0.80 & \wedge & 0.20 \\ 0.40 & \wedge & 0.50 \\ 0.80 & \wedge & 0.55 \\ 0.40 & \wedge & 0.75 \\ 0.40 & \wedge & 0.95 \\ 0.40 & \wedge & 1.00 \end{bmatrix}$$

and, therefore,

$$\underset{X}{\int} h_2(x) \circ g(.) \quad = \quad 0.55.$$

In this case, although candidate 2 has satisfied the age and teaching criteria to high degrees, poor research ability has led to the low evaluation.

c) Candidate 1 is selected for the post due to a higher evaluation.

d) Let us suppose that a secretary has misplaced information about the candidates' research abilities. However, due to time limitations, a candidate must be selected. Additionally, the members of the selection committee have given very low default values for research ability [$h_1(x_3) = h_2(x_3) = 0.2$: the other values remain the same]. Hence, we obtain

$$\underset{X}{\int} h_1(x) \circ g(.) \quad = \quad \text{Max} \begin{bmatrix} 0.00 & \wedge & 0.00 \\ 0.40 & \wedge & 0.10 \\ 0.70 & \wedge & 0.20 \\ 0.20 & \wedge & 0.50 \\ 0.40 & \wedge & 0.55 \\ 0.20 & \wedge & 0.75 \\ 0.20 & \wedge & 0.95 \\ 0.20 & \wedge & 1.00 \end{bmatrix}$$

$$= \quad \text{Max}(0.00,\ 0.10,\ 0.20,\ 0.20,\ 0.40,\ 0.20,\ 0.20,\ 0.20)$$

or

$$\underset{X}{\int} h_1(x) \circ g(.) \quad = \quad 0.40,$$

for candidate 1; and

$$\oint_X h_2(x) \circ g(.) \quad = \quad \text{Max} \begin{bmatrix} 0.00 & \wedge & 0.00 \\ 1.00 & \wedge & 0.10 \\ 0.80 & \wedge & 0.20 \\ 0.20 & \wedge & 0.50 \\ 0.80 & \wedge & 0.55 \\ 0.20 & \wedge & 0.75 \\ 0.20 & \wedge & 0.95 \\ 0.20 & \wedge & 1.00 \end{bmatrix}$$

$$= \quad \text{Max}(0.00,\ 0.10,\ 0.20,\ 0.55,\ 0.20,\ 0.20,\ 0.20,\ 0.20)$$

or

$$\oint_X h_2(x) \circ g(.) \quad = \quad 0.55$$

for candidate 2.

Candidate 2 is selected because of higher satisfaction levels of the criteria for which information is available. Note that the highest possible evaluation is 0.55, since the two criteria are not too important in the selection procedure.

CONCLUDING REMARKS

The underlying principle of probability, be it aleatory or epistemic, is monotonicity. Aleatory probability, because of its relation to frequencies, is necessarily additive. In contrast, epistemic probability, which is purely a feature of the human mind, need not be so restricted. We feel that non-additive monotonic measures lend mathematical formality to the study of subjectivity and its existence in human reasoning.

In our attempt to introduce human subjectivity into mechanistic decision making, we have broken down the decision strategy into two separate parts. One is the intrinsic importance that propositions carry, and the other, is the extent to which a manifestation satisfies each proposition. Non-additive measures provide numerical weights or levels of importance to sets of propositions. These values are assumed to be known *a priori*. On the other hand, the *truth value* of each proposition would depend on the manifestation, and is the result of observation.

The Sugeno Integral combines these two quantities non-linearly, and results in an overall evaluation of the manifestation. This functional, which is also monotonic, has excellent intuitive features, and in our opinion, effectively approximates a human evaluation. Additionally, when probability measures are employed, the Sugeno Integral is close to the probability expectation value (or, Lebesgue Integral). Thus, the Sugeno evaluation may be considered to be a

subjective expectation value in applications where monotonicity is used to grasp the concept of subjectivity.

The field of Expert Systems is one of the most active and exciting areas of research in AI. A high performance expert system must incorporate human subjectivity in its decision making. For this reason, we feel that monotonic measures and the Sugeno Integral, defined on these measures, could find applications in the design and construction of efficient expert systems.

REFERENCES

1. Arnauld, Antoine, and Nicole, Pierre, "la Logique, ou l'art de penser", Paris (1662). [c.f. Shafer (1978)]

2. Banon, G., "Distinction entre Plusiers Sous-Ensembles de Mesures Floues", Proc. Colloq. Int. Theorie Appl. Sous-Ensembles Flous, Marseilles (1978).

3. Bayes, Thomas, "An essay towards solving a problem in the doctrine of chance", Phil. Trans. Roy. Soc. London for 1763, London (1764). Reprinted in Biometrika, 45, 293-315 (1958).

4. Bernoulli, Jacob, "Ars Conjectandi", Basel (1713). [c.f. Shafer (1978); and Hacking (1975)]

5. Brakel, J. van, "Some Remarks on the Prehistory of the Concept of Statistical Probability", Archive for History of Exact Sciences, 16, 119-136 (1976).

6. Condorcet, Marie Jean, "Essai sur l'application de l'analyse a la probabilite des decisions rendues a la pluralite des voix", Paris (1785). [c.f. Shafer (1978)]

7. De Moivre, Abraham, "De Mensura Sortis, seu, de Probabilitate Eventum in Ludis a Casu Fortuito Pendentibus", Phil. Trans. Roy. Soc. London, 27, 213-264 (1711). [c.f. Shafer (1978)]

8. Dempster, A. P., "Upper and lower probabilities induced by a multivalued mapping", Ann. Math. Statist., 38, 325-339 (1967).

9. Dubois, D. and Prade, H., "Fuzzy Sets and Systems: Theory and Applications", 125-146, Academic Press, New York (1980).

10. Dubois, D. and Prade, H., "Unfair Coins and Necessity Measures: Towards a Possibilistic Interpretation of Histograms", Fuzzy Sets and Systems, 10, 15-20 (1983).

11. Hacking, Ian, "The Emergence of Probability", Cambridge Univ. Press, London (1975).

12. Kolmogorov, A. N., "Foundations of the Theory of Probability", Berlin (1933).

13. Lambert, J. H., "Neues Organon", Leipzig (1764). [c.f. Shafer (1978); and Hacking (1975)]

14. Laplace, P. S., "Oeuvres Completes", 14 vols., Paris (1878-1912). [c.f. Shafer (1978); and Hacking (1975)]

15. Montmort, P. R., "Essay d'analyse sur les jeux de hazard", Paris (1708). [c.f. Shafer (1978)]

16. Sekita, Y., and Tabata, Y., "A consideration on identifying fuzzy measures", paper presented at XXIII Int. Mtg. of the Institute of Management Sciences, Athens (1977).

17. Shafer, G., "A Mathematical Theory of Evidence", Princeton Univ. Press, Princeton, N.J. (1976).

18. Shafer, G., "Non-additive probabilities in the work of Bernoulli and Lambert", Archive for History of Exact Sciences, 19, 309-370 (1978).

19. Sugeno, M., "Theory of Fuzzy Integrals and Its Applications", Ph.D. Thesis, Tokyo Institute of Technology, Tokyo (1974).

20. Terano, T., and Sugeno, M., "Fuzzy Sets and Their Applications to Cognitive and Decision Processes" (L. A. Zadeh, K. S. Fu, K. Tanaka, and M. Shimura, Eds.), 151-170, Academic Press, New York (1975).

21. Wierzchon, S. T., "An Algorithm for the Identification of Fuzzy Measures", Fuzzy Sets and Systems, 9, 69-78 (1983).

22. Zadeh, L. A., "Fuzzy Sets as a Basis for a Theory of Possibility", Fuzzy Sets and Systems, 1, 3-28 (1978).

CHAPTER IV

THE SUGENO INTEGRAL IN THE PRODUCTION RULE FORMALISM

The technical issues of acquiring, representing, and using knowledge are important problems in Knowledge Engineering, and in the design of expert systems. *Production Rules* (or *IF-THEN rules*) are a popular approach for representing and manipulating domain facts and heuristics in expert systems. Implemented in *Rule-based*, or *Production Systems*, these rules are natural to human strategies of problem solving and decision making.

In this chapter, we focus on production systems, and go on to propose a methodology for evaluating the premises of production rules based on the concepts of *monotonic measures* and the *Sugeno Integral*. The methodology essentially deals with the combination of evidence in the production rule formalism, and provides an excellent foundation for expressing, representing, and coping with the subjectivity that is often introduced into human evaluations. Additionally, the methodology offers a convenient framework for the representation and treatment of ignorance, and the conservatism that is seen in evaluations made in its presence. Finally, we extend the formulation so as to admit multilevel reasoning.

PRODUCTION SYSTEMS

Production Systems were first proposed by Post (1943) as a formal mechanism for combinatorial decision problems. Newell and Simon (1972) later incorporated them in their models of human cognition, and since then, the methodology has undergone substantial theoretical development which has led to its extensive use in several AI programs. A production system is a modular knowledge representation scheme that has been found to be useful as a mechanism for controlling the interaction between declarative and procedural knowledge (Barr and Feigenbaum, 1981). This has made it a popular approach for representing both facts and heuristics of domain knowledge in an expert system. Production systems are founded on a notion of *condition-action* (or situation-action, or IF-THEN) rules known as *production rules*, or simply, *productions*.

All production systems have three basic components: a set of production rules that forms the *rule base* for problem solving, a context or *data base* that helps in evaluating the rules, and, an *interpreter* that controls the system's activity by using the rules to mainpulate the data base.

A production rule is a conditional statement written in the form

If this condition holds, then this action is appropriate.

This scheme represents both logical implication

A implies B

as well as causality

A causes B,

and is a convenient methodology for dealing with humanistic reasoning. For example, the oft-used thumb-rule in football (Barr and Feigenbaum, 1981)

Always punt on fourth down with long yardage required,

may be translated to the production rule

IF it is fourth down AND long yardage is required THEN punt.

The IF part of a production rule (also called the condition part or left-hand side) stipulates the conditions that must be satisfied if the production rule is to be applicable. In general, this is a complex conditional statement comprised of simple or unitary propositions joined by AND and OR connectors. The THEN part (or action part, or right-hand side) defines the action to be taken. A production whose condition part is satisfied can *fire*, that is, have its action part executed by the interpreter. The invocation of many rules in a production system can be viewed as a chained sequence of *modus ponens* actions. This is a data-driven (or bottom-up) strategy, and it is possible to vary the methodology to obtain a goal-driven (top-down, or backward) scheme. Here, the elements of the left-hand side are interpreted to be the goals obtained by the successful matching of elements from the right-hand side. In this case the rules *unwind.* Thus, the same set of rules can be used in two different ways, with characteristically different control structures, and possibly, behavior. In some instances it may be feasible to attempt a solution to a problem by moving bi-directionally, that is, both forward and backward simultaneously (Nilsson, 1980).

The data base is the focus of attention of production rules. Also known as the short-term memory buffer, it contains the state variables, the facts and assertions about the world. Before a production rule can fire, each element of its condition part must be present in the context data structure. This may be a simple list, a large array, or even a medium-size buffer with an internal structure of its own. But whatever the organization of the data base, it is the sole storage medium for all the state variables of the system, and all information must go there. Moreover, the store is universally accessible to every rule in the system, so that anything located there is potentially detectable by any rule. This is termed as the unity of data and control store.

The interpreter, which is the source of much of the variation found among different systems, controls the system's activity by adjusting the sequence of application of the rules. The simplest interpreter operates in a *select-execute* loop, in which a rule applicable to the current state of the data base is chosen and then executed. The action results in a modified data base, and the select phase begins again. This alternation of selection and execution is an essential element of production system architecture, and is responsible for a very fundamental feature. Since a rule is selected for execution on the basis of the total contents of the data base, a complete re-evaluation of the control state of the system is performed at each cycle. This is distinctly different fron procedurally-oriented approaches, and production systems are potentially sensitive to any changes in the entire environment.

There are several approaches to the selection procedure. Data-driven approaches utilize variations of a left-hand scan, in which each left-hand side is evaluated in turn. In such designs, conflict resolution is an important consideration. Some systems resolve conflicts by stopping their scan at the first successful evaluation. However, once this is done, the question of where to start the next scan remains to be solved.

MYCIN (Shortliffe, 1976), which is goal-directed, uses a right-hand scan. Given a sub-goal, it examines all rules whose actions conclude something about the sub-goal. Evaluation of the first right-hand side is undertaken, and if any clause in it refers to a fact not already present in the data base, a generalized version of the fact becomes the new sub-goal, and the process recurs. Since MYCIN is designed to deal with judgmental knowledge (implemented in meta-rules and certainty factors), it does not stop after the first success. Instead, it evaluates all possible rules and estimates the certainties of their conclusions. Thus, the use of meta-level knowledge is seen to aid in conflict resolution.

Interpreter architecture strongly influences the overall efficiency of a production system, and the specific domain of application has an important bearing on the design of the interpreter. Interested readers are referred to Davis and King (1977), and, Nilsson (1980) for excellent treatments of interpreters for production systems.

The use of a production system methodology has several advantages. Production rules offer a modular representation of knowledge that is easily accessed and modified. The rules do not call each other, and communicate only through the data base. In the process, interaction

between the rules themselves is kept to a minimum, and each rule is almost an independent piece of knowledge. Rules may, therefore, be added, deleted, or modified independently. This modularity in knowledge representation is useful as a scheme for systems designed to approach competence in an incremental fashion. Encoding domain knowledge in the form of rules reduces the *entropy* within the system and imposes a uniform structure on the knowledge within the rule base. This facilitates human understanding of the problem solving process. Better synthetic understanding is also achieved by the machine itself, and this translates into a more efficient explanation facility. An added advantage is that production systems are natural to human understanding of problem solving. Most experts, when asked about their knowledge, find it convenient to express it in the form of production rules. This allows for easier filtering of domain knowledge during design, and also permits knowledge acquisition by the system.

There are, however, significant disadvantages in production system formulations. The uniformity and modularity of knowledge representation give rise to large overheads in problem solving. Often, each action is performed by a *select-execute* cycle, and all information must be communicated to the context data structure. This creates inefficiencies in program execution. It is not possible to program sequencies of actions that may be required in certain applications. Larger steps, or *leaps* in reasoning are, therefore, never permitted. Another disadvantage is the opacity of control flow in problem solving, due to the fact that production systems are distinctly different from procedural approaches,

and are, therefore, not easily represented in algorithmic form.

Production systems easily solve problems in some domains, but are rather inappropriate for others. Production rules capture effectively knowledge that is diffuse, consisting of many facts and rules of thumb. Most classification and diagnosis problems fall in this category. On the other hand, the methodology fails when applied to areas in which a few tenets embody much of the domain knowledge. For this reason, production rules are not able to capture knowledge in concise fields, such as mathematics and physics. The complexity of control flow is also important in determining whether production systems are appropriate. They are suitable for modeling processes which can be represented as a set of independent actions, but are awkward for complex, parallel processes with dependent sub-processes, for which procedural approaches may be better suited. An important feature of production system architecture is that the data base is completely separated from the interpreter, and the methodology makes no prior assumptions about the way facts are employed. Thus, production systems are appropriate when domain knowledge can be separated from the way it is to be used. Fields in which representation and control are merged are better treated procedurally.

An excellent overview of production systems is presented by Davis and King (1977). Other pertinent references are Nilsson (1980), Barr and Feigenbaum (1981,1982), Cohen and Feigenbaum (1982), and Gevarter (1983).

THE SUGENO INTEGRAL AS THE BASIS FOR EVALUATING PRODUCTION RULES

In this section we propose a general formulation for representing and evaluating production rules. The methodology provides a convenient framework for approximate reasoning in the presence of ambiguity as well as partial ignorance. We start by examining production rule formulations in two well-known expert systems.

A production rule is a statement cast in the form

IF *condition* THEN *action*.

One of the rules that the expert system, R1 (McDermott, 1982), summons while configuring VAX computers is:

IF	the most current active context is assigning a power supply
AND	an SBI module of any type has been put in a cabinet
AND	the position it occupies in the cabinet (its nexus) is known
AND	there is space available for a power supply for that nexus
AND	there is an available power supply
THEN	put the power supply in the cabinet in the available space.

R1 operates this rule by matching the conditions to the current situation. Each proposition or assertion in the premise is matched with corresponding elements in the data base. The rule is fired only if all the conditions that make up the premise are satisfied. The propositions in the premise are *crisp* - answered by *yes* or *no*, *true* or *false*, or, *1* or *0*.

This is the simplest form of a production rule. Mathematically, the evaluation of the premise is performed by a minimum operator, i.e.,

$$\text{Min}(h(x_1), h(x_2), h(x_3), h(x_4), h(x_5)) \tag{4.1}$$

where

$$h(x_i) \in \{0,1\}, \tag{4.2}$$

represents the truth value of proposition x_i, such that

$$h(x_i) = 1, \text{ if proposition } x_i \text{ is true (satisfied),}$$

and

$$h(x_i) = 0, \text{ if proposition } x_i \text{ is false (not satisfied).}$$

If

$$\underset{i}{\text{Min}}(h(x_i)) = 1, \tag{4.3}$$

the condition part of the rule is satisfied, and the rule can be fired. On the other hand, if

$$\underset{i}{\text{Min}}(h(x_i)) = 0, \tag{4.4}$$

at least one of the propositions is false, and the premise of the rule is, therefore, not satisfied. In general, the left-hand side of the rule could be a complex conditional statement comprised of unitary or atomic propositions linked by AND and OR connectors. Evaluation would then involve the use of minimum (AND) and maximum (OR) operators.

Judgment plays an important role in clinical diagnosis, and MYCIN introduces a new level of complexity in its rule base. The rule

```
IF      the infection which requires therapy is meningitis
   AND  the patient has evidence of a serious skin infection
   AND  organisms were not seen on the strain of the culture
   AND  the type of infection is bacterial
THEN    there is evidence that the organism (other than those
        seen on cultures or smears) which might be causing the
        infection is
        staphylococcus-coagpus  (0.75)
        streptococcus  (0.50)
```

is used by MYCIN to diagnose and prescribe therapy for bacterial infectious diseases. The left-hand side still consists of simple assertions, but uncertainty is introduced into the right-hand side. For the rule presented above, the evidence cited in the premise provides degrees of confidence (rule certainty factors, CFs) of 0.75 and 0.50 for staphylococcus-coagpus and streptococcus, respectively. The CFs are measures of association between the premise and action clauses for each rule in MYCIN's rule base. They assume that all the antecedents are known with absolute certainty. If the rules' antecedents are not determined to be completely true, the certainty factors for the conclusions are reduced accordingly.

The premise of each rule in MYCIN is a *boolean* combination of one or more clauses. It is always a conjunction of clauses, but may contain complex conjunctions and disjunctions nested within each clause (Instead of writing a rule whose premise would be a disjunction of clauses, a separate rule is written for each clause). Each clause is represented in the form of a 4-tuple,

(<predicate function> <object> <attribute> <value>).

Thus, to cope with situations in the real-world, medical facts are represented in the form of 4-tuples corresponding to an atomic formula with a numeric truth value between -1.0 and 1.0. A value of -1.0 implies complete confidence that the proposition is false, while a value of 1.0 represents complete confidence in its truth. A proposition is given a value of 0 if there is no evidence for its truth or falsehood. This allows for the combination of evidence both, in favor of and against

the same hypothesis. For example,

(TYPE INFECTION BACTERIAL 0.70)

is interpreted as *the type of infection is bacterial is known with a certainty of 0.70.* Hence, depending on the evidence, MYCIN permits its propositions to have varying levels of truth and falsehood.

As in predicate calculus, the rules of inference provide a basis for combining well-formed-formulas and truth values. MYCIN's model of approximate reasoning employs a unique calculus for combining evidence. When the premise of a rule is evaluated, each predicate returns a number between -1.0 and 1.0. The AND connector necessitates a minimization of the arguments, while an OR connector requires that a maximization be performed (recall that a rule may have nested OR conditions in its premise). Thus, evaluation of a premise results in a numerical value between -1.0 and 1.0. For a rule whose premise evaluation does not lie within the empirically determined interval (-0.2,0.2), the conclusion is made with a certainty that is the product of the premise evaluation and the certainty factor of the rule.

Approximate reasoning is a process by which a possible imprecise conclusion is deduced from a collection of imprecise premises. In fact, it is the ability to reason in qualitative, imprecise terms that distinguishes human intelligence from machine intelligence. Imprecision is a feature of the real-world, and, whether it is caused by uncertainty, ambiguity, or even, ignorance, human beings display an ability to reason in its presence. The conclusions,

of course, would then be less than completely certain. In the framework of production rules, imprecision gives rise to ambiguous premises which are rarely completely satisfied. A suitable approach for dealing with imprecision is to *soften* production rules so that even partial satisfaction of their premises could lead to some action being taken by the interpreter.

The first step in creating *soft* production rules is to allow the propositions that make up the premises to take truth values between truth and falsehood. Several conventional expert systems provide multivalued truth values to their propositions. For example, MYCIN's propositions are assigned values in the range -1.0 to 1.0, as suggested by the clinical evidence. As the following example illustrates, human knowledge often consists of facts that can only be stated imprecisely. It is this feature that admits propositions that can have varying degrees of truth.

<u>Example 4.2.1</u>

Most people who are familiar with the game of football have a general idea of the qualities a good running back should possess. A running back is usually well-built, but not too bulky. More importantly, he should be very quick, and must have excellent ball-handling ability. This knowledge can be represented in the following production rule;

```
IF        a man is well-built, but not too bulky
    AND   he is very quick
    AND   he possesses excellent ball-handling ability
THEN      he would make a good running back.
```

The premise consists of three propositions, pertaining to the attributes: build, speed, and ball-handling ability. Note that the propositions are judgmental, and appear to be ambiguous. However, this is a feature of knowledge about a running back, and even expert football scouts would agree that the rule is reasonable. Indeed, the same is also true in many other domains. Experts often find it difficult to express their knowledge exactly. The facts and definitions are usually very qualitative, and the evaluations are purely judgmental. Yet, an expert is able to provide excellent results. Perhaps, this is because the expert possesses a deeper, more correct conception of domain knowledge.

Consider the proposition concerning a running back's speed,

he is very quick.

When an expert (himself, a well-known running back) was asked for additional details, his response was

"...he should be able to run the 40-yard dash in 4.7 seconds, or less. And, the faster, the better."

This is the expert's deeper perspective. He knows what to look for, and how to rate what he sees. Thus, depending on his speed, a candidate could satisfy the proposition to a greater or lesser extent. This degree of satisfaction is the truth value of the proposition (with reference to the candidate). A possible relation between the time, t, taken to run 40 yards, and the truth value of the proposition, $h(x_2)$, is:

$$h(x_2) = 1, \quad \text{if} \quad t < 4.5 \text{ seconds}$$

$$h(x_2) = 1.0 - 0.25(t - 4.5), \quad \text{if} \quad t \in [4.5, 4.9] \text{ seconds}$$

and

$$h(x_2) = 0, \quad \text{if} \quad t > 4.9 \text{ seconds.}$$

The presence of linguistic concepts in the propositions gives rise to ambiguity, and necessitates multivalued truth levels. Perhaps, the essential difference between an interested spectator and an expert scout is in the fact that the latter has a better notion of these linguistic concepts that qualify the attributes. It is this better notion that we call *expertise*. Hence, on observation of a candidate, the expert could provide better truth values to the propositions.

As Example 4.2.1 demonstrates, ambiguities in facts and definitions arises from their specification in qualitative, linguistic terms (see, e.g., Zadeh, 1975a, 1975b, 1975c; concerning the use of linguistic variables in approximate reasoning). Another characteristic of human reasoning is that one is rarely absolutely confident of the truth or falsehood of propositions. To accomodate the ambiguity and uncertainty inherent in many areas of human expertise, it is plausible to permit propositions to take multivalued levels of truth. This is a general formulation that would always embrace conventional *crisp* propositions.

A production rule can be represented in the general form:

```
IF      proposition 1
   AND  proposition 2
   AND  proposition 3
    .        .
    .        .
THEN    action.
```

The left-hand side of the rule presented above, is a compound condition comprised of unitary or atomic propositions linked by AND connectors. We assume that premises having OR connectors can always be decomposed into two or more production rules of the general form, with the same action, or right-hand side. For example, the production

```
IF      proposition 1
   AND (proposition 2  OR  proposition 3)
THEN    action
```

can be broken down to two rules having the same action,

```
IF      proposition 1
   AND  proposition 2
THEN    action
```

and

```
IF      proposition 1
   AND  proposition 3
THEN    action.
```

The right-hand side of a rule in this formulation is not restricted to a single action. More generally, it could consist of a set of actions to be performed once the conditions are satisfied.

Since the premise of a production rule in the present framework is restricted to a compound statement consisting of AND-connected, soft propositions, for purposes of evaluation, it is convenient to represent it as a set of propositions. For example, the rule

```
IF      proposition 1
   AND  proposition 2
    .        .
    .        .
   AND  proposition n
THEN    action
```

would have its premise, X, given by

$$X = \{x_1, x_2, \ldots, x_n\} \tag{4.5}$$

where x_i represents the i-th proposition in the premise, X. The production rule can, therefore, be written as

IF X THEN action.

Depending on the observation or manifestation, truth values,

$$h(x_i) \in [0,1], \tag{4.6}$$

are provided to each proposition, $x_i \in X$. Many conventional expert systems (R1 and MYCIN, included) perform evaluations of their premises (containing AND-connected propositions) using the minimum operator. The premise evaluation, $E_m(X)$, for the rule is given by

$$E_m(X) = \underset{x_i \in X}{\mathrm{Min}}(h(x_i)). \tag{4.7}$$

There are two important points to be noted concerning this operation. First, the evaluation is very *pessimistic*, since the premise evaluation is the lowest truth value from among the propositions contained in the premise set. Secondly, the evaluation using the minimum operator is also feasible for *crisp* propositions, for which

$$h(x_i) \in \{0,1\}. \tag{4.2}$$

Human experts introduce considerable subjectivity into their decision making. Some extent of subjectivity is incorporated into production rules due to the specification of premises in terms of *soft* propositions. However, also important is the fact that in performing evaluations, human beings are inclined to weigh and balance the evidence. We often have prior conceptions concerning

the relative weights of propositions. These notions play important roles in our analyses.

Example 4.2.2

Knowledge pertaining to the selection of a good running back (Example 4.2.1) is represented in the production rule:

```
IF        a man is well-built, but not too bulky
   AND    he is very quick
   AND    he possesses excellent ball-handling ability
THEN      he would make a good running back.
```

The propositions that make up the premise of the rule deal with three criteria: build, speed, and ball-handling ability. We felt that, while all three criteria are relevant in selecting a running back, some qualities could, possibly, be more important than others. So we went back to our old friend, the expert, for his opinion. His answers were most enlightening.

"...Most important, perhaps, is his speed. A good running back is almost always an excellent sprinter. Take Herschel Walker, for example..."

"...But then, he should be able to hold on to the ball without fumbling. So ball-handling ability would be almost as important. Well, perhaps, slightly lower on the scale..."

"...Build is surely not as important as the other two criteria. But we can't have someone too small or too big, playing running back. In any case, it would be unlikely for someone rather big to be able to run real fast..."

The comments provided above illustrate that knowledge about the relative weights of propositions does form a sizeable chunk of expertise. Additionally, they are usually known *a priori*.

Monotonicity is a fundamental feature that is inherent in human evaluative strategies. The principle of monotonicity is illustrated by the adage

Given more, we feel at least as good, or even better.

Since a premise of a production rule is written as a set, it is convenient to employ measures of sets to represent the magnitudes of importance that groups of propositions (or subsets of the premise set) carry. The principle of monotonicity is used as the basis for the definition of these measures. Monotonic measures (or fuzzy measures, see, e.g., Sugeno, 1974) provide a plausible framework for dealing with information concerning the relative weights of propositions.

Definition 4.2.1 (Sugeno, 1974)

A *monotonic* (or fuzzy) *measure*, g, defined on a finite reference set, X, is a function from the σ-algebra, P(X), or power set of X to the interval [0,1], which has the following properties;

i) $g(\emptyset) = 0$, $g(X) = 1$. (Boundedness and Non-negativity) (4.8a)

ii) $\forall Q_1, Q_2 \in P(X)$, if $Q_1 \subseteq Q_2$,

$g(Q_1) \leq g(Q_2)$. (Monotonicity) (4.8b)

The premise of a production rule is the finite reference set, X, in Definition 4.2.1, and any subset Q of X would then be a subset of propositions from the premise. The measure, g(Q), is taken to represent the collective importance that the criteria (propositions) contained in

set Q contribute toward the evaluation of the premise. The complete premise, X, represents the totality of evidence that can be used in the evaluation, and is, therefore, assigned a measure of one, i.e.,

$$g(X) = 1. \tag{4.8a}$$

On the other hand, the null set, $\emptyset$, contains no propositions, and, therefore, contributes no information towards the evaluation; thus,

$$g(\emptyset) = 0. \tag{4.8a}$$

The second property in Definition 4.2.1, which gives monotonicity to the measures, provides the necessary intuitive framework. If Q_1 is a subset of Q_2, the set Q_2 contains at least one more criterion (proposition) than set Q_1. It would, therefore, carry at least as much weight as (if not more than) set Q_1 in the evaluation. Thus, we write

$$g(Q_1) \leqslant g(Q_2). \tag{4.8b}$$

The broad definition of monotonicity includes *additive* as well as *non-additive* features. A monotonic measure which is subject to the additional constraint

$$g(Q_1 \cup Q_2) = g(Q_1) + g(Q_2), \text{ for } Q_1, Q_2 \varepsilon\ P(X) \text{ and } Q_1 \cap Q_2 = \emptyset, \tag{4.9}$$

is additive. This restriction implies that the combined measure is exactly equal to the sum of the individual measures of sets Q_1 and Q_2. Or, in our framework, the total information content or importance of a group of criteria is exactly equal to the sum of the contributions of the individual criteria. The principle of additivity (Equation 4.9) is used to define probability measures in stochastic

theory, and finds its rationale in the frequentative interpretation of randomness.

The phenomenon of human subjectivity in the combination of evidence is purely an epistemic concern. We are dealing with *a priori* notions that may have nothing in common with the paradigm of chance. It is implausible to suggest that these notions always follow the axiom of additivity. A more reasonable approach would be to allow for general non-additivity in the domain of evidence combination. Additivity of measures is restricted to special cases. Hence, instead of Equation 4.9, non-additive monotonic measures are subject to the additional constraint,

$$g(Q_1 \cup Q_2) \gtreqless g(Q_1) + g(Q_2), \text{ for } Q_1,\ Q_2 \varepsilon\ P(X) \text{ and } Q_1 \cap Q_2 = \emptyset. \tag{4.10}$$

The following example demonstrates the utility of monotonic measures in knowledge representation.

<u>Example 4.2.3</u>

In Examples 4.2.1 and 4.2.2, the premise, X, of the production rule contains three propositions, i.e.,

$$X = \{x_1,\ x_2,\ x_3\}$$

where x_1 concerns the person's build, x_2 concerns the person's speed, and, x_3 concerns the person's ball-handling ability. The production rule can, therefore, be written in the form

IF X THEN Y (i.e., the person would make a good running back)

or

IF $\{x_1,\ x_2,\ x_3\}$ THEN Y.

Corresponding to this formulation we could define sub-rules of the original production. The premises for these sub-rules are subsets of the original premise, and, in general, 2^n separate sub-rules can be generated from a premise consisting of n propositions. The premise in our example has three propositions, and the following eight sub-rules result.

i) IF $\emptyset$ THEN Y

ii) IF $\{x_1\}$ THEN Y

iii) IF $\{x_2\}$ THEN Y

iv) IF $\{x_3\}$ THEN Y

v) IF $\{x_1, x_2\}$ THEN Y

vi) IF $\{x_1, x_3\}$ THEN Y

vii) IF $\{x_2, x_3\}$ THEN Y

and

viii) IF $\{x_1, x_2, x_3\}$ THEN Y.

Each sub-rule can be used independently in an evaluation. However, depending on the importance of the propositions contained in the premise of a sub-rule, a greater or smaller weight is accorded to it. This would later manifest itself in the evaluation. The weights of importance of the premises of sub-rules represent the meta-level knowledge that is inherent in human expertise. An expert would know exactly how important each subset of his original premise is, and the effect it would have on his evaluation.

Monotonic measures provide a useful methodology for representing meta-level knowledge. The premise, X, corresponds to the totality of

evidence that can be used in an evaluation. Hence

$$g(X) = 1.$$

On the other hand, the empty premise in the sub-rule

IF $\emptyset$ THEN Y

bears no information, and thus,

$$g(\emptyset) = 0.$$

The expert's remarks in Example 4.2.2 indicate that although the three attributes of build, speed, and ball-handling ability are relevant, all of them are not equally important for the task of selecting a running back. Speed is the most preferred attribute, while it is not absolutely essential that the person must satisfy the build criterion. Based on these aspects, the following measures for sets of criteria from the original premise (or, degrees of importance for premises of sub-rules) have been provided by the expert.

$$g(\emptyset) = 0$$

$$g(\{x_1\}) = 0.10$$

$$g(\{x_2\}) = 0.25$$

$$g(\{x_3\}) = 0.20$$

$$g(\{x_1, x_2\}) = 0.50$$

$$g(\{x_1, x_3\}) = 0.40$$

$$g(\{x_2, x_3\}) = 0.60$$

and

$$g(\{x_1, x_2, x_3\}) = 1.$$

The measures are seen to be monotonic and non-additive.

One important point remains to be noted. It is almost a fact that

an obese person cannot run very fast. One might, therefore, argue that it is incorrect to treat build and speed as independent criteria, as the theory of monotonic measures dictates. We concede that the relation between build and speed does exist. However, we are dealing with epistemic concepts, and the subjectivity that enters in their combination is, perhaps, too complex to model precisely. In order to admit mathematical treatment of these concepts, the present approach assumes that each proposition is an independent piece of knowledge. Interactions, such as the one between build and speed, are implicitly dealt with in the procedure for formulating the measure for their combination.

Before venturing any further, it is useful to review the ideas we expounded to this point. We have seen that domain-specific knowledge encoded in the production rule formalism offers manifold advantages. However, imprecision is a feature of the real-world, and allowances must be made for reasoning in its presence. The specification of the premise of a production rule in terms of *soft* propositions that can take *shades* of truth, is the first step toward reasoning in the presence of imprecision. An expert tends to have a clear conception of the propositions that make up his expertise. He knows what the propositions mean, and how to evaluate them depending on what he sees. Hence, the task of a knowledge engineer is to find out exactly what the soft propositions imply. He should also define them in a manner that would enable a computer to assign accurate truth values depending on the manifestation.

Human experts also have prior notions about the relative weights of propositions in a premise. Some propositions are more important than others, and this deeper information is taken into account when the evidence is weighed and balanced. Monotonic measures have been shown to possess properties that come in useful for representing meta-knowledge of this type. Perhaps, a person could be considered to be an expert because he has clearer notions concerning the relative importances of criteria. If this is true, Knowledge Engineering requires us to glean this information from the expert so that the most accurate measures of importance are obtained.

In the present attempt to introduce human subjectivity into mechanistic decision making, the decision strategy has been decomposed into two distinct parts. One is the intrinsic importance that each proposition (and each group of propositions) carries in an evaluation, and the other is the extent to which the propositions are satisfied once an observation is made. A functional is needed to combine these two aspects so that a *mean* evaluation of the premise is achieved. The Sugeno Integral (or, Fuzzy Integral, see, e.g., Sugeno, 1974) is one such functional that also has excellent intuitive justification.

Definition 4.2.2 (Sugeno, 1974)

For a finite reference set X, let (X, P(X), g) be a monotonic measure space, and let $h:X \to [0,1]$ be a function defined on X. The *Sugeno Integral* over any finite set $Q \in P(X)$ is given by

$$\int_Q h(x)\circ g(.) = \underset{F \varepsilon P(X)}{\mathrm{Max}} \left(\underset{x_i \varepsilon F}{\mathrm{Min}} (h(x_i)) \wedge g(Q \cap F) \right). \quad (4.11)$$

The Sugeno Integral combines the function, h, with the monotonic measure, g, in a non-linear fashion that may appear to be rather enigmatic; however, it possesses useful mathematical properties. The Sugeno Integral may be interpreted as representing a mean or expected value in applications where monotonic measures are used to grasp human subjectivity. The integral is itself monotonic, and is analogous to the Lebesgue Integral (a functional defined on additive measures) that finds applications in the theory of probability. In fact, Sugeno (1974) has shown that if the integral is defined on a probability measure space (i.e., the measure, g, follows the axioms of probability), the value is close to the probability expectation value. In addition to these mathematical properties, the Sugeno Integral has excellent intuitive features that render it useful for the evaluation of premises of production rules. For the time being, let us confine ourselves to the Sugeno Integral defined over the finite reference set X, which is written as

$$\int_X h(x)\circ g(.) = \underset{F \varepsilon P(X)}{\mathrm{Max}} \left(\underset{x_i \varepsilon F}{\mathrm{Min}} (h(x_i)) \wedge g(X \cap F) \right)$$

$$= \underset{F \varepsilon P(X)}{\mathrm{Max}} \left(\underset{x_i \varepsilon F}{\mathrm{Min}} (h(x_i)) \wedge g(F) \right). \quad (4.12)$$

Recall that in the present framework, the premise of a production rule is the reference set, X. Each element of this set is a proposition or criterion, and the premise is a collection of criteria on the basis

of which the manifestation must be evaluated. Intuitively, this corresponds to a complete picture, while a subset F of X, is a partial or incomplete view of the overall picture. Information about the importance of the view F is specified by the monotonic measure, g(F). The term, $h(x_i)$, is the truth value or level of satisfaction of the criterion, x_i, with reference to the manifestation. The operation

$$\min_{x_i \varepsilon F} (h(x_i)) \qquad (4.13)$$

in Equation (4.12) may be interpreted as providing the pessimistic, or most secure level of satisfaction that the manifestation offers when examined from the point of view of the criteria contained in the view F.

In its attempt to provide a mean evaluation, the Sugeno Integral combines the most secure level of satisfaction obtained from a view with the relative importance of that particular view. This is done for all possible views [F εP(X)] of the overall picture, X. These values are then further combined, resulting in the mean evaluation. Two specific operations are performed. First, the most secure level of satisfaction is combined with the relative importance of the corresponding view using the minimum operator. The value obtained is no greater than the importance of the view, g(F); and this serves to limit the evaluation provided by a view to a value no greater than its importance. The second combination, which involves the maximum operator, selects the best evaluation from among those provided by

all possible views. This is a common tendency in human judgments. We look at an object from many different angles or aspects. The angle that *strikes* us in terms of satisfaction as well as importance plays a major role in our analyses.

Example 4.2.4

In previous examples, we have seen that knowledge pertaining to the selection of a running back is embodied in the production rule

IF a man is well-built, but not too bulky (x_1)
AND he is very quick (x_2)
AND he has excellent ball-handling ability (x_3)
THEN he would make a good running back (Y)

or

IF X THEN Y

where

$$X = \{x_1, x_2, x_3\}.$$

Let us suppose that an expert football scout has given the following evaluations to a walk-on candidate for the football team;

$$h(x_1) = 0.35$$

$$h(x_2) = 0.45$$

and

$$h(x_3) = 0.55.$$

Or, the candidate has been evaluated as satisfying the criterion of build to a degree of 0.35, speed to a degree of 0.45, and, ball-handling ability to a degree of 0.55.

a) The conventional scheme involves the use of a minimum operator.

The expert systems, R1 and MYCIN use this methodology [see, Equation (4.7)], and the premise evaluation is given by

$$
\begin{aligned}
E_m(X) &= \min_{x_i \varepsilon X}(h(x_i)) \\
&= \mathrm{Min}(h(x_1), h(x_2), h(x_3)) \\
&= \mathrm{Min}(0.35, 0.45, 0.55) \\
&= 0.35.
\end{aligned}
$$

The blanket use of the minimum operator results in a very pessimistic evaluation, since the candidate is evaluated to be only as good as his worst quality indicates. More importantly, the deeper, yet pertinent, information concerning the relative importances of the three criteria has been neglected in the evaluation.

b) In Example 4.2.3, we have seen that the original production can be broken down into 2^3 (= 8) separate sub-rules, given by

IF F THEN Y

where

$$F \subseteq X,$$

i.e.,

$$F \;\varepsilon P(X) = \{\emptyset, \{x_1\}, \{x_2\}, \{x_3\}, \{x_1, x_2\}, \{x_1, x_3\}, \{x_2, x_3\}, X\}.$$

Each sub-rule is a specific piece of knowledge which could be used independently in deciding whether the candidate would make a good running back. However, depending on the criteria contained in its premise, the sub-rule carries a varying level of importance as an independent evaluation. In terms of the intuitive picture discussed previously, the premise of a sub-rule corresponds to a view of the overall picture. The view may be incomplete, in which case, the

resulting evaluation is not absolutely certain. The measures, $g(F)$, where $F \varepsilon P(X)$, carry information pertaining to the importance of each view. Our expert has provided the following values;

$$g(\emptyset) = 0$$

$$g(\{x_1\}) = 0.10$$

$$g(\{x_2\}) = 0.25$$

$$g(\{x_3\}) = 0.20$$

$$g(\{x_1, x_2\}) = 0.50$$

$$g(\{x_1, x_3\}) = 0.40$$

$$g(\{x_2, x_3\}) = 0.60$$

and

$$g(\{x_1, x_2, x_3\}) = 1.$$

The Sugeno Integral takes each sub-rule (or, view), and calculates the minimum or most secure level of satisfaction for its premise. Thus, the operation

$$\underset{x_i \varepsilon F}{\mathrm{Min}}(h(x_i))$$

gives rise to

$$0 \quad \text{for} \quad F = \emptyset,$$

$$0.35 \quad \text{for} \quad F = \{x_1\},$$

$$0.45 \quad \text{for} \quad F = \{x_2\},$$

$$0.55 \quad \text{for} \quad F = \{x_3\},$$

$$\mathrm{Min}(0.35, 0.45) = 0.35 \quad \text{for} \quad F = \{x_1, x_2\},$$

$$\mathrm{Min}(0.35, 0.55) = 0.35 \quad \text{for} \quad F = \{x_1, x_3\},$$

$$\mathrm{Min}(0.45, 0.55) = 0.45 \quad \text{for} \quad F = \{x_2, x_3\},$$

and

$$\text{Min}(0.35, 0.45, 0.55) = 0.35 \quad \text{for} \quad F = \{x_1, x_2, x_3\}.$$

Next, the most secure level of satisfaction and the importance of the premise of each sub-rule are combined. The evaluation provided by any sub-rule is limited to a value no greater than its importance, and the operation

$$\underset{x_i \varepsilon F}{\text{Min}}(h(x_i)) \wedge g(F)$$

yields

$$0 \wedge 0 = 0 \quad \text{for} \quad F = \emptyset,$$

$$0.35 \wedge 0.10 = 0.10 \quad \text{for} \quad F = \{x_1\},$$

$$0.45 \wedge 0.25 = 0.25 \quad \text{for} \quad F = \{x_2\},$$

$$0.55 \wedge 0.20 = 0.20 \quad \text{for} \quad F = \{x_3\},$$

$$0.35 \wedge 0.50 = 0.35 \quad \text{for} \quad F = \{x_1, x_2\},$$

$$0.35 \wedge 0.40 = 0.35 \quad \text{for} \quad F = \{x_1, x_3\},$$

$$0.45 \wedge 0.60 = 0.45 \quad \text{for} \quad F = \{x_2, x_3\},$$

and

$$0.35 \wedge 1.00 = 0.35 \quad \text{for} \quad F = \{x_1, x_2, x_3\}.$$

Finally, the Sugeno Integral takes the best (highest) evaluation from among those provided by all the sub-rules. The premise evaluation is given by

$$\begin{aligned} E_s(X) &= \int_X h(x) \circ g(.) \\ &= \text{Max}(0, 0.10, 0.25, 0.20, 0.35, 0.35, 0.45, 0.35) \\ &= 0.45. \end{aligned}$$

Note that the value obtained using the Sugeno Integral [$E_s(X) = 0.45$], is substantially higher than the evaluation obtained using the minimum operator in part (a) [$E_m(X) = 0.35$]. This is because we have been able to take into account the information concerning the relative weights of propositions in the evaluation. The criteria of speed and ball-handling ability are considered to be rather important when taken together [$g(\{x_2, x_3\}) = 0.60$]. Our candidate has satisfied these criteria, collectively, to a value of 0.45; this has resulted in the higher evaluation.

Monotonic measures have been introduced in the context of production rules so that the *a priori* notions that human experts have about the relative weights of propositions in premises can be represented and dealt with effectively. The Sugeno Integral, defined on these monotonic measures, has been shown to have a reasonable intuitive justification as a means for evaluating premises. On the other hand, the evaluation employed by R1 and MYCIN does not take into account the importances of propositions, and has been seen to be rather pessimistic. We employ the concept of a *vacuous belief function* (Shafer, 1976) to demonstrate an interesting relationship between the Sugeno Integral and the conventional procedure of evaluation, using the minimum operator.

Definition 4.2.3 (Shafer, 1976)

A *vacuous belief function*, g_v, is a monotonic measure defined on

a finite reference set, X, that satisfies the condition

$$\forall Q \subset X, \quad g_v(Q) = 0. \tag{4.14}$$

Or, alternatively,

$$g_v(Q) = 1, \quad \text{if } Q = X, \tag{4.15a}$$

and

$$g_v(Q) = 0, \quad \text{if } Q \text{ is any other subset of } X. \tag{4.15b}$$

Let us examine the Sugeno Integral defined on the measure space of a vacuous belief function, $(X,\ P(X),\ g_v)$. The Sugeno Integral over a finite reference set, X, is given by

$$\begin{aligned} \oint_X h(x) \bullet g(.) &= \underset{F \varepsilon P(X)}{\mathrm{Max}} \left(\underset{x_i \varepsilon F}{\mathrm{Min}} (h(x_i)) \wedge g(X \cap F) \right) \\ &= \underset{F \varepsilon P(X)}{\mathrm{Max}} \left(\underset{x_i \varepsilon F}{\mathrm{Min}} (h(x_i)) \wedge g(F) \right). \end{aligned} \tag{4.12}$$

Thus,

$$\oint_X h(x) \circ g_v(.) = \underset{F \varepsilon P(X)}{\mathrm{Max}} \left(\underset{x_i \varepsilon F}{\mathrm{Min}} (h(x_i)) \wedge g_v(F) \right). \tag{4.16}$$

Since

$$h(x_i) \ \varepsilon \ [0,1] \tag{4.6}$$

and

$$\begin{aligned} g_v(F) &= 1, \quad \text{if } F = X && (4.15a) \\ &= 0, \quad \text{if } F \subset X, && (4.15b) \end{aligned}$$

for $F \subset X$, we obtain

$$\begin{aligned} \underset{x_i \varepsilon F}{\mathrm{Min}}(h(x_i)) \wedge g_v(F) &= \underset{x_i \varepsilon F}{\mathrm{Min}}(h(x_i)) \wedge 0 \\ &= 0 \end{aligned} \tag{4.17}$$

or, more specifically,

$$\underset{\substack{x_i \varepsilon F \\ F \subset X}}{\mathrm{Min}}(h(x_i)) \wedge g_v(F) = 0. \tag{4.18}$$

On substituting Equation (4.18) into Equation (4.16), we obtain

$$\oint_X h(x) \bullet g_v(.) = \underset{F \varepsilon P(X)}{\text{Max}} \left(\underset{x_i \varepsilon F}{\text{Min}} (h(x_i)) \wedge g_v(F) \right)$$

$$= \text{Max} \left(\underset{\substack{x_i \varepsilon F \\ F \subset X}}{(\text{Min}}(h(x_i)) \wedge g_v(F)), \underset{\substack{x_i \varepsilon F \\ F = X}}{(\text{Min}}(h(x_i)) \wedge g_v(F)) \right)$$

$$= \text{Max} \left(0, 0, \ldots, 0, \underset{x_i \varepsilon X}{(\text{Min}}(h(x_i)) \wedge g_v(X)) \right)$$

$$= \underset{x_i \varepsilon X}{\text{Min}} (h(x_i)) \wedge g_v(X). \tag{4.19}$$

Additionally, since

$$h(x_i) \ \varepsilon \ [0,1] \tag{4.6}$$

and

$$g_v(X) = 1, \tag{4.15a}$$

Equation (4.19) reduces to

$$\oint_X h(x) \bullet g_v(.) = \underset{x_i \varepsilon X}{\text{Min}} (h(x_i)) \wedge 1$$

$$= \underset{x_i \varepsilon X}{\text{Min}} (h(x_i)). \tag{4.20}$$

Equation (4.20) leads to the following lemma.

Lemma 4.2.1

The Sugeno Integral over a finite reference set, X, and defined on the measure space of a vacuous belief function, $(X, P(X), g_v)$, reduces to the conventional evaluation using the minimum operator. Or, the following relation holds;

$$\oint_X h(x) \bullet g_v(.) = \underset{x_i \varepsilon X}{\text{Min}} (h(x_i)). \tag{4.20}$$

Lemma 4.2.1 provides a significant relationship between the Sugeno Integral and the conventional evaluation using the minimum operator. In the present framework, monotonic measures are employed to convey information about the relative weights of propositions in a premise. The vacuous belief function is a special case of a monotonic measure that corresponds to *abject ignorance* about the relative weights. This measure gives no importance to all evaluations based on incomplete views (or, sub-rules). Only the complete premise has a non-zero level of importance, and, since it corresponds to the totality of evidence, it is given a weight of one. This implies that all the criteria must be used in the evaluation of a premise, and a partial or incomplete set of criteria will not suffice.

We have stated that the conventional scheme of evaluation is pessimistic. However, the pessimism involved now has added significance. The fact that the Sugeno Integral under the conditions of vacuous belief is equivalent to the minimum operator, implies that the pessimism that is a feature of the latter could be considered to be the result of ignorance. Almost always in real-life, when we know more about something, we are more certain about our results. Ignorance, or knowing nothing about the relative weights is implicit in the definition of vacuous belief, and the Sugeno Integral under these conditions, should correspond to the lowest possible evaluation. This fact is, indeed, true, and follows from Lemma 4.2.1. In general, for constant h(x),

$$\oint_X h(x) \bullet g(.) \geq \oint_X h(x) \bullet g_V(.) \qquad (4.21a)$$

or, by Lemma 4.2.1,

$$\oint_X h(x) \circ g(\cdot) \geq \underset{x_i \varepsilon X}{\mathrm{Min}}(h(x_i)) \tag{4.21b}$$

where $g(.)$ is any monotonic measure; and, $g_v(.)$ is the vacuous belief function.

The development of expert systems that can reason in the presence of ignorance is the focus of ongoing research. Yet, as Duda and Shortliffe (1983) point out, "questions about how a program should reason in the presence of ignorance, or how it can even recognize the limits of its knowledge, are largely unanswered". It is, perhaps, impossible to develop mechanisms that can reason in the presence of complete ignorance, and our attention, therefore, is confined to situations in which ignorance is not total. The definition of the vacuous belief function, and its use in the Sugeno Integral represents a step in this direction. A vacuous belief function entails the absence of the meta-level knowledge that is used to weigh and balance the evidence prior to making a decision. The Sugeno Integral in this situation is the lowest bound or most pessimistic value. The low value is the consequence of ignorance. Observation of human behavior also shows that ignorance is often associated with conservatism and pessimism, and these lead to lower levels of certainty being assigned to the resultant decisions. The present treatment of ignorance is based on this fact.

We have so far confined ourselves to the Sugeno Integral evaluated over the reference set, X. The premise of a production rule is the

reference set, and the Sugeno Integral has been used to obtain a *mean* evaluation over the entire premise, given by

$$\begin{aligned} E_s(X) &= \int_X h(x) \circ g(.) \\ &= \max_{F \in P(X)} \left(\min_{x_i \in F} (h(x_i)) \wedge g(X \cap F) \right) \\ &= \max_{F \in P(X)} \left(\min_{x_i \in F} (h(x_i)) \wedge g(F) \right). \end{aligned} \tag{4.22}$$

The evaluation, $E_s(X)$, has been shown to have excellent intuitive features. Since the premise is represented in terms of a set of criteria, there are several views or aspects on the basis of which an object may be examined. The Sugeno Integral combines the most secure level of satisfaction obtained from each view, with the relative importance of that particular view. Each view, therefore, provides a partial evaluation, and the integral selects the best evaluation from among all possible views. As stated previously, the Sugeno Integral appears to model the human tendency, whereby, we look at an object from many different angles. The angle that strikes us the most plays a major role in the final analyses. Since the evaluation is performed over the entire premise of the production rule, the Sugeno Integral considered up to now attempts to examine the object from all possible views. The power set, P(X), lists the maximum number of views that can be used.

We have mentioned that ignorance, or not knowing enough, leads to conservative evaluations. In terms of the intuitive picture, however, it is suitable to consider conservatism in evaluations

within the paradigm of ignorance as arising out of an inability to examine an object from all possible aspects. Complete knowledge creates no such limitation; in this situation, it is possible to look at an object from all angles.

To illustrate this point, suppose that we are required to make a judgment concerning the girth of a tree-trunk. We would first attempt to walk around the tree, and only then, would we make our evaluations. On the other hand, if a barrier prevented us from circumnavigating the tree, the analyses would be incomplete, and the resulting judgments would not be very certain. In the present treatment, we view ignorance as the allegoric barrier that prevents circumspection. The list of possible views is no longer the power set, but instead, is a proper subset of P(X). And, the greater the ignorance, the fewer the views, and hence, the smaller the subset of the power set over which information is integrated.

The term

$$E_s(X) \quad = \quad ⨍_X h(x) \circ g(\cdot) \tag{4.22}$$

represents an evaluation of the entire premise. The evaluation is made on the basis of all the criteria, or all possible views, and hence, the upper bound on $E_s(X)$ is one [This is because in Equation (4.22) $g(F) = 1$, for $F = X$; if $h(x_i) = 1$ for all $x_i \in X$, $E_s(X) = 1$]. Consider now, the Sugeno Integral evaluated over a proper subset Q of the reference set, X. Since one or more criteria present in the premise are not contained in set Q, the corresponding evaluation is

incomplete, and is, at best, a partial evaluation, given by (see, Definition 4.2.2)

$$E_s(Q) = \oint_Q h(x) \bullet g(.)$$

$$= \underset{F \varepsilon P(X)}{\mathrm{Max}} \left(\underset{x_i \varepsilon F}{\mathrm{Min}} (h(x_i)) \wedge g(Q \cap F) \right). \qquad (4.23)$$

This corresponds to a situation of partial ignorance in the present framework. Note that the measure, g, on the right-hand side of Equation (4.23) has the set, $Q \cap F$, as its argument. Although F can be any element of the power set of X [$F \varepsilon P(X)$], the argument, $Q \cap F$, limits the number of views on the basis of which the evaluation is made. The new list of views is given by P(Q), where P(Q) is the power set of set Q. Additionally, the Sugeno Integral is restricted to a value no greater than g(Q). In general, since

$$Q \subseteq X, \qquad (4.24)$$

due to the definition of monotonic measures (Definition 4.2.1),

$$g(Q) \leq g(X) \qquad (4.8b)$$

and we obtain

$$E_s(Q) \leq E_s(X) \qquad (4.25)$$

where

$$E_s(Q) = \oint_Q h(x) \bullet g(.) \qquad (4.23)$$

and

$$E_s(X) = \oint_X h(x) \bullet g(.). \qquad (4.22)$$

Equation (4.25) demonstrates monotonicity for the Sugeno Integral.

In the present treatment of evaluations made in the presence of ignorance, the set Q, over which the integration is performed, is a subset of the premise set, X. The complete set of criteria is not taken into account in the evaluation, and, therefore, the result can never be absolutely certain. Monotonicity of the Sugeno Integral is a key point in the treatment of reasoning in the presence of ignorance. It can be seen that the smaller the set Q, over which the integration is performed (or, the fewer the criteria considered), the lower the value of the Sugeno Integral. This monotonicity is convenient for dealing with ignorance, and also effectively models the conservatism that arises from this situation. Note that we adopt a methodology by which an evaluation made in the paradigm of total ignorance is always assigned a value of zero, or *false*, i.e.,

$$\begin{aligned} E_s(\emptyset) &= \fint_{\emptyset} h(x) \bullet g(.) \\ &= \underset{F \varepsilon P(X)}{\text{Max}} \left(\underset{x_i \varepsilon F}{\text{Min}} (h(x_i)) \wedge g(\emptyset \cap F) \right) \\ &= \underset{F \varepsilon P(X)}{\text{Max}} \left(\underset{x_i \varepsilon F}{\text{Min}} (h(x_i)) \wedge g(\emptyset) \right) \\ &= 0. \end{aligned} \tag{4.26}$$

[Recall that due to the definition of monotonic measures (see Definition 4.2.1), $g(\emptyset) = 0$.]

The following example demonstrates the applicability of the Sugeno Integral.

Example 4.2.5

Continuing with the problem of selecting a running back, recall that the production rule is given by

IF a man is well-built, but not too bulky (x_1)
AND he is very quick (x_2)
AND he possesses excellent ball-handling ability (x_3)
THEN he would make a good running back (Y).

This is equivalent to the rule

IF X THEN Y

where

$$X = \{x_1, x_2, x_3\}.$$

Let us suppose that for some reason it is not possible to evaluate the candidate's ball-handling ability. This could occur in a situation in which information about the extent to which the candidate satisfies this attribute has not been provided. Hence, only the criteria of build and speed are taken into account, and the production rule reduces to

IF a man is well-built, but not too bulky (x_1)
AND he is very quick (x_2)
THEN he would make a good running back (Y).

The premise of this production rule is a subset of the original premise, X; it is given by

$$Q = \{x_1, x_2\},$$

corresponding to the sub-rule

IF Q THEN Y.

In Examples 4.2.3 and 4.2.4, we have seen that the relative importances of groups of criteria are

$g(\emptyset) = 0$

$g(\{x_1\}) = 0.10$

$g(\{x_2\}) = 0.25$

$g(\{x_3\}) = 0.20$

$g(\{x_1, x_2\}) = 0.50$

$g(\{x_1, x_3\}) = 0.40$

$g(\{x_2, x_3\}) = 0.60$

and

$g(\{x_1, x_2, x_3\}) = 1.$

These values correspond to the eight different views on the basis of which the original premise is evaluated. Since the measure

$g(\{x_1, x_2, x_3\}) = 1,$

the Sugeno Integral evaluation of the entire premise $E_s(X)$, has an upper bound of one. On the other hand, the Sugeno Integral defined over the set Q is employed to evaluate the incomplete sub-rule, and

$$\begin{aligned} E_s(Q) &= \oint_Q h(x) \circ g(.) \\ &= \max_{F \in P(X)} \left(\min_{x_i \in F} (h(x_i)) \wedge g(Q \cap F) \right). \end{aligned} \qquad (4.23)$$

The argument, $Q \cap F$, for the measure, g, restricts the evaluation to just four different views. This is because

$g(Q \cap \emptyset) = g(\emptyset)$, for $F = \emptyset$

$g(Q \cap \{x_1\}) = g(\{x_1\})$, for $F = \{x_1\}$

$g(Q \cap \{x_2\}) = g(\{x_2\})$, for $F = \{x_2\}$

$$g(Q \cap \{x_3\}) = g(\emptyset), \quad \text{for} \quad F = \{x_3\}$$

$$g(Q \cap \{x_1, x_2\}) = g(\{x_1, x_2\}), \quad \text{for} \quad F = \{x_1, x_2\}$$

$$g(Q \cap \{x_1, x_3\}) = g(\{x_1\}), \quad \text{for} \quad F = \{x_1, x_3\}$$

$$g(Q \cap \{x_2, x_3\}) = g(\{x_2\}), \quad \text{for} \quad F = \{x_2, x_3\}$$

and

$$g(Q \cap \{x_1, x_2, x_3\}) = g(\{x_1, x_2\}), \quad \text{for} \quad F = \{x_1, x_2, x_3\}.$$

The measures corresponding to these four views are given by

$$g(\emptyset) = 0$$

$$g(\{x_1\}) = 0.10$$

$$g(\{x_2\}) = 0.25$$

and

$$g(\{x_1, x_2\}) = 0.50.$$

Due to the operations involved in the Sugeno Integral [Equation (4.23)], we can see that

$$E_s(Q) \leq g(Q \cap \{x_1, x_2, x_3\}) = g(\{x_1, x_2\}) = 0.50.$$

In Example 4.2.4, the degrees to which the walk-on candidate has satisfied the three propositions are

$$h(x_1) = 0.35$$

$$h(x_2) = 0.45$$

and

$$h(x_3) = 0.55$$

The present example, however, does not take into account the criterion of ball-handling ability. This corresponds to ignorance about the attribute, and, a suitable default value is

$h(x_3) = 0.$

The other truth values, $h(x_1)$ and $h(x_2)$, are unchanged. In assigning the value of zero to the proposition, we assume that a worst-case evaluation is given to a criterion that is not considered. This smacks of pessimism; however, it is a common trend in human judgment to give a conservative estimate concerning something that is unknown. Thus

$$
\begin{aligned}
E_s(Q) &= \underset{Q}{⨍}\, h(x) \circ g(.) \\
&= \underset{F \varepsilon P(X)}{\text{Max}} \left(\underset{x_i \varepsilon F}{\text{Min}} \, (h(x_i)) \wedge g(Q \cap F) \right) \\
&= \text{Max} \begin{bmatrix} 0.00 \wedge 0, & \text{for} \quad F = \emptyset \\ 0.35 \wedge 0.10, & \text{for} \quad F = \{x_1\} \\ 0.45 \wedge 0.25, & \text{for} \quad F = \{x_2\} \\ 0.00 \wedge 0, & \text{for} \quad F = \{x_3\} \\ 0.35 \wedge 0.50, & \text{for} \quad F = \{x_1, x_2\} \\ 0.00 \wedge 0.10, & \text{for} \quad F = \{x_1, x_3\} \\ 0.00 \wedge 0.25, & \text{for} \quad F = \{x_2, x_3\} \\ 0.00 \wedge 0.50, & \text{for} \quad F = \{x_1, x_2, x_3\} \end{bmatrix} \\
&= \text{Max}(0, 0.10, 0.25, 0, 0.35, 0, 0, 0) \\
&= 0.35.
\end{aligned}
$$

Note that the value, $E_s(Q) = 0.35$, obtained above, is significantly lower than the evaluation of the premise in Example 4.2.4 [$E_s(X) = 0.45$]. This lower value is expected.

An interesting point to note is that the evaluation of the

incomplete premise, Q, under the conditions of vacuous belief is given by

$$E_V(Q) = \int_Q h(x) \circ g_V(.)$$

$$= 0. \tag{4.27}$$

This is because the vacuous belief function places absolutely no importance on all incomplete views. Indeed, the evaluation is performed strictly on an *all or nothing* basis.

THE SUGENO INTEGRAL IN MULTILEVEL REASONING

In the preceding sections, we have seen that production rules offer a modular representation of knowledge that captures the essence of human expertise. The methodology proposed by us employs the Sugeno Integral to evaluate the premises of *soft* production rules which are comprised of AND-connected propositions. This provides a convenient framework for approximate reasoning in the presence of ambiguity as well as partial ignorance. Additionally, the conventional evaluation that uses the minimum operator has been shown to be a special case of the present methodology. In this section, we extend the formulation to admit multilevel reasoning.

A production rule connects a set of conditions with one or more actions that are relevant once the conditions have been satisfied. This represents a single deductive step. During the operation of a production system, the interpreter summons the short-term memory or data base, each time it executes a rule. If the premise of the rule is satisfied, the rule is fired, and the corresponding actions modify the data base. The interpreter proceeds to control the activity of the system by adjusting the sequence of application of the rules. Rules are selected and executed according to a predetermined sequence, and each time, the data base is accessed. This elicits a chain of reasoning, in which, a series of production rules appear to be linked. The action parts of production rules at one level form the premises of production rules at another level. Rules nested in this manner

represent larger deductive steps, or, *deductive leaps*.

As an example, consider the following production rules;

IF $X^{(1)}$ THEN $Y^{(1)}$

IF $X^{(2)}$ THEN $Y^{(2)}$

and

IF $X^{(3)}$ THEN $Y^{(3)}$

where $X^{(1)}$, $X^{(2)}$, and $X^{(3)}$ are the premises of three different rules. Let us assume that the interpreter selects, executes, and then fires each rule in turn. The actions, $Y^{(1)}$, $Y^{(2)}$, and $Y^{(3)}$, then proceed to modify the data base.

Suppose that there exists a fourth rule which uses these actions as conditions in its premise. This rule could be written as

IF $X^{(4)}$ THEN $Y^{(4)}$

or

IF $Y^{(1)}$

AND $Y^{(2)}$

AND $Y^{(3)}$

THEN $Y^{(4)}$.

The scheduling, execution, and firing of the four rules in sequence, represent two unitary deductive steps, or a deductive leap spanning two levels.

At this stage, it is of interest to examine the reasoning methodology that MYCIN (Shortliffe, 1976) employs. Suppose that the following productions are present in MYCIN's rule base;

Rule 1: IF $X^{(1)}$ THEN $Y^{(1)}$ [$CF^{(1)}$]

Rule 2: IF ($X^{(2a)}$ OR $X^{(2b)}$) THEN $Y^{(2)}$ [$CF^{(2)}$]

Rule 3: IF $X^{(3)}$ THEN $Y^{(3)}$ [$CF^{(3)}$]

and an additional rule that uses the actions, $Y^{(1)}$, $Y^{(2)}$, and $Y^{(3)}$, as propositions in its premise. This rule could be specified as

Rule 4: IF ($Y^{(1)}$ AND $Y^{(2)}$ AND $Y^{(3)}$) THEN $Y^{(4)}$

or,

Rule 4: IF $X^{(4)}$ THEN $Y^{(4)}$ [$CF^{(4)}$].

Note that $X^{(i)}$ represents an AND-connected set of propositions, and the term, $CF^{(i)}$, is the certainty factor for the i-th rule.

Let us suppose that data concerning the premises of Rules 1, 2, and 3, are already present in the data base. Additionally, let us assume that the action of Rule 4, $Y^{(4)}$, is the prescribed goal. MYCIN is a goal-directed system, and the process of reasoning starts by searching the rule base for rules whose actions conclude something about the goal. This points to Rule 4. Next, the premise of Rule 4 is examined. The clauses, $Y^{(1)}$, $Y^{(2)}$, and $Y^{(3)}$, which comprise the premise, $X^{(4)}$, become the new sug-goals, and MYCIN proceeds to search its rule base for more information. This procedure comes up with Rules 1, 2, and 3. A tree is generated, and using truth values

$$h \in [-1.0, 1.0], \qquad (4.28)$$

for the propositions in Rules 1, 2, and 3, the certainty or truth value of the goal, $Y^{(4)}$, is established. The methodology is as follows:

First, the premises of Rules 1, 2, and 3, are evaluated by resorting to conventional maximum (OR), and minimum (AND) operators;

the results are given by

$$E_m(X^{(1)}) = \min_{x_j \in X^{(1)}} (h(x_j)) \tag{4.29}$$

$$E_m(X^{(2)}) = \text{Max} \left(E_m(X^{(2a)}),\ E_m(X^{(2b)}) \right)$$

$$= \text{Max} \left(\min_{x_j \in X^{(2a)}} (h(x_j)),\ \min_{x_j \in X^{(2b)}} (h(x_j)) \right) \tag{4.30}$$

and

$$E_m(X^{(3)}) = \min_{x_j \in X^{(3)}} (h(x_j)). \tag{4.31}$$

Next, the certainties of the actions are obtained by multiplying the premise evaluations with the corresponding rule certainty factors. Note that if a premise evaluation, E_m, lies in the interval (-0.2,0.2), there is insufficient evidence pointing to truth, or falsehood, and the certainty of the corresponding action is equated to zero. Otherwise,

$$h(Y^{(1)}) = CF^{(1)} \cdot E_m(X^{(1)}) \tag{4.32}$$

$$h(Y^{(2)}) = CF^{(2)} \cdot E_m(X^{(2)}) \tag{4.33}$$

and

$$h(Y^{(3)}) = CF^{(3)} \cdot E_m(X^{(3)}). \tag{4.34}$$

The premise of Rule 4 can now be evaluated. The result is given by

$$E_m(X^{(4)}) = \min_{x_j \in X^{(4)}} (h(x_j))$$

$$= \text{Min} \left(h(Y^{(1)}),\ h(Y^{(2)}),\ h(Y^{(3)}) \right). \tag{4.35}$$

Finally, the certainty or truth value of the goal $Y^{(4)}$ is estimated. This is specified by

$$h(Y^{(4)}) = CF^{(4)} \cdot E_m(X^{(4)}) \tag{4.36a}$$

$$= 0, \text{ if } E_m(X^{(4)}) \in (-0.2, 0.2). \tag{4.36b}$$

In the present methodology, premises of production rules are restricted to those consisting of AND-connected propositions. This is a convenient approach that gives rise to uniformity and modularity within the rule base. Additionally, in our opinion, human experts find it easier to comprehend and specify rule certainty factors for rules that are made up of conjunctions of simple propositions. The OR connector is quite ambiguous, and logical ORs nested within premises normally complicate matters. More importantly, it will be seen that the decomposition of complex premises containing disjunctions, into two or more rules having the same action, is performed with no loss in generality. In fact, the methodology involving the decomposition and subsequent evaluation of complex premises is the more general of the two methodologies.

Let us consider the four rules examined previously. Rules 1, 3, and 4 satisfy the restriction placed on the premises of production rules in the present treatment. However, Rule 2 has an OR connector in its premise, and must be decomposed into two simple rules having the same action. The new rule base is shown below.

Rule 1: IF $X^{(1)}$ THEN $Y^{(1)}$ [$CF^{(1)}$]

Rule 2a: IF $X^{(2a)}$ THEN $Y^{(2)}$ [$CF^{(2a)}$]

Rule 2b: IF $X^{(2b)}$ THEN $Y^{(2)}$ [$CF^{(2b)}$]

Rule 3: IF $X^{(3)}$ THEN $Y^{(3)}$ [$CF^{(3)}$]

and

Rule 4: IF $X^{(4)}$ THEN $Y^{(4)}$ [$CF^{(4)}$].

For this modified rule base, let us examine the evaluation of the certainty of the goal $Y^{(4)}$. The premises of Rules 1, 2a, 2b, and 3 offer the following evaluations;

$$E_m(X^{(1)}) = \operatorname*{Min}_{x_j \varepsilon X^{(1)}} (h(x_j)) \tag{4.37}$$

$$E_m(X^{(2a)}) = \operatorname*{Min}_{x_j \varepsilon X^{(2a)}} (h(x_j)) \tag{4.38}$$

$$E_m(X^{(2b)}) = \operatorname*{Min}_{x_j \varepsilon X^{(2b)}} (h(x_j)) \tag{4.39}$$

and

$$E_m(X^{(3)}) = \operatorname*{Min}_{x_j \varepsilon X^{(3)}} (h(x_j)). \tag{4.40}$$

There are now two distinct ways in which we can proceed. Either, we can essentially combine the evaluations of Rules 2a and 2b, as MYCIN does, or we can go on to estimate the certainties of the sub-goals, $Y^{(1)}$, $Y^{(2)}$, and $Y^{(3)}$. Following the latter procedure, we obtain

$$h(Y^{(1)}) = CF^{(1)} \cdot E_m(X^{(1)}) \tag{4.41}$$

$$h(Y^{(2)})_{2a} = CF^{(2a)} \cdot E_m(X^{(2a)}), \quad \text{from Rule 2a} \tag{4.42}$$

$$h(Y^{(2)})_{2b} = CF^{(2b)} \cdot E_m(X^{(2b)}), \quad \text{from Rule 2b} \tag{4.43}$$

and

$$h(Y^{(3)}) \quad = \quad CF^{(3)} \cdot E_m(X^{(3)}). \tag{4.44}$$

We now have two estimates of the certainty or truth value of the sub-goal $Y^{(2)}$ [Equations (4.42) and (4.43)]. Since Rules 2a and 2b are two distinct pieces of knowledge that point toward the same conclusion, it is feasible to select the certainty of $Y^{(2)}$ using the maximum operator. Thus,

$$\begin{aligned} h(Y^{(2)}) \quad &= \quad \text{Max} \; (\; h(Y^{(2)})_{2a}, \; h(Y^{(2)})_{2b} \;) \\ &= \quad \text{Max} \; (\; CF^{(2a)} \cdot E_m(X^{(2a)}), \; CF^{(2b)} \cdot E_m(X^{(2b)}) \;). \end{aligned} \tag{4.45}$$

Note that this result is equivalent to the certainty estimated in the previous case [see Equation (4.33)], only if

$$CF^{(2)} \quad = \quad CF^{(2a)} \quad = \quad CF^{(2b)}. \tag{4.46}$$

This is because, if Equation (4.46) holds, we obtain

$$\begin{aligned} h(Y^{(2)}) \quad &= \quad \text{Max} \; (\; CF^{(2a)} \cdot E_m(X^{(2a)}), \; CF^{(2b)} \cdot E_m(X^{(2b)}) \;) \\ &= \quad CF^{(2)} \cdot \text{Max} \; (\; E_m(X^{(2a)}), \; E_m(X^{(2b)}) \;). \end{aligned} \tag{4.47}$$

Through the use of Equation (4.30), we can write

$$h(Y^{(2)}) \quad = \quad CF^{(2)} \cdot E_m(X^{(2)}), \tag{4.33}$$

which is the estimate obtained prior to the modification of the rule base.

In general, the two certainty factors do not have to be equal, and, therefore, the results provided by the two approaches do indeed differ. We adopt the modified-rule-base approach which is more general, and is also justified from the point of view of convenience. Having obtained the certainties of the sub-goals, $Y^{(1)}$, $Y^{(2)}$, and $Y^{(3)}$; the estimation of $h(Y^{(4)})$ proceeds as before. Thus,

$$E_m(X^{(4)}) = \underset{x_j \varepsilon X^{(4)}}{\text{Min}} (h(x_j))$$

$$= \text{Min} (h(Y^{(1)}), h(Y^{(2)}), h(Y^{(3)})) \qquad (4.48)$$

and

$$h(Y^{(4)}) = CF^{(4)} \cdot E_m(X^{(4)}). \qquad (4.49)$$

[Note that Equations (4.48) and (4.49) are the same as Equations (4.35) and (4.36) respectively.]

It is now a simple matter to incorporate the Sugeno Integral in multilevel reasoning. As already stated, the present formulation deals with production rules whose premises consist of AND-connected soft propositions. Furthermore, the truth values or degrees of certainty, h, that are assigned to propositions are restricted to the [0,1] interval, i.e.,

$$h \ \varepsilon \ [0,1] \qquad (4.6)$$

where "0" corresponds to falsehood, and "1" entails truth or complete satisfaction. In the conventional scheme, the premise evaluation is given by

$$E_m(X) = \underset{x_j \varepsilon X}{\text{Min}} (h(x_j)). \qquad (4.7)$$

We have seen that this evaluation is equivalent to the Sugeno Integral defined on a vacuous belief function measure space, $(X, P(X), g_v)$ [This fact follows from Lemma 4.2.1.], i.e.,

$$E_m(X) = \underset{x_j \varepsilon X}{\text{Min}} (h(x_j))$$

$$= \oint_X h(x) \circ g_v(.). \qquad (4.20)$$

Monotonic measures are employed to convey meta-level information about the relative weights that groups of propositions carry in the evaluation of a premise. The vacuous belief function is a special monotonic measure that corresponds to an absence of this deeper information. In general, for any monotonic measure, g, the value of the Sugeno Integral is never less than the value obtained by using the minimum operator (or, the Sugeno Integral defined under the conditions of vacuous belief). This fact is stated mathematically as

$$E_s(X) \geqslant E_m(X) = E_v(X) \tag{4.50}$$

where

$$E_s(X) = \oint_X h(x) \bullet g(.) \tag{4.22}$$

$$E_m(X) = \min_{x_j \varepsilon X} (h(x_j)) \tag{4.7}$$

and

$$E_v(X) = \oint_X h(x) \bullet g_v(.) . \tag{4.51}$$

The higher value for $E_s(X)$ in Equation (4.50) is expected, and arises from the use of additional or meta-level knowledge in the evaluation. Since the evaluation can cope with additional information, it is more general than the conventional minimum operator; and is employed in the present methodology to evaluate the premises of production rules.

Before we can specify a *generic* production rule, one more point remains to be considered. After a premise has been evaluated, it is necessary to establish the certainty (or truth value) of the

corresponding action. Recall that MYCIN's model of approximate reasoning employs a certainty factor, CF, for each action in a production rule, where

$$CF \in [0,1]. \tag{4.52}$$

A certainty factor that is close to one represents a strong link between the conditions and action of the production rule. In other words, this implies that the action has a high degree of certainty when the premise is satisfied. For a rule i, whose premise evaluation does not lie within the interval, (-0.2, 0.2), MYCIN's conclusion is made with a certainty, $h(Y^{(i)})$, that is the product of the premise evaluation, $E_m(X^{(i)})$, and the corresponding rule certainty factor, $CF^{(i)}$. Thus,

$$h \in [-1.0,\ 1.0] \tag{4.28}$$

and

$$h(Y^{(i)}) = 0, \quad \text{if } E_m(X^{(i)}) \in [-0.2,\ 0.2] \tag{4.53a}$$

$$h(Y^{(i)}) = CF^{(i)} \cdot E_m(X^{(i)}), \quad \text{otherwise.} \tag{4.53b}$$

In general, however, a *certainty function*, $\Psi^{(i)}$, defined by

$$\Psi^{(i)}: E_m(X^{(i)}) \ [\text{or, } E_s(X^{(i)})] \rightarrow [0,1], \tag{4.54}$$

is suitable for expressing the strength of the link between the conditions and action of the i-th production rule. Human experts usually find it quite convenient to specify their conceptions in linguistic terms. The notion of *linguistic variables* (see, e.g., Zadeh, 1975a, 1975b, 1975c) provides a rationale for the representation of linguistic concepts in terms of the certainty

functions, $\Psi^{(i)}$. These certainty functions are similar to the certainty factors in MYCIN; however, the certainty or truth value of the action, $Y^{(i)}$, is given by

$$h(Y^{(i)}) = \Psi^{(i)}(E_s(X^{(i)})). \tag{4.55}$$

We may now express a *generic* production rule in the form

$$\begin{array}{llll} \text{Rule i:} & \text{IF} & X^{(i)} & [\ g^{(i)}(.)\ ] \\ & \text{THEN} & Y_1^{(i)} & [\ \Psi_1^{(i)}\ ] \\ & & Y_2^{(i)} & [\ \Psi_2^{(i)}\ ] \\ & & \vdots & \quad \vdots \\ & & Y_{m_i}^{(i)} & [\ \Psi_{m_i}^{(i)}\ ] \end{array} \tag{4.56}$$

where $X^{(i)}$ is the premise of Rule i, which consists of n_i AND-propositions, and is written as

$$X^{(i)} = \{x_1^{(i)}, x_2^{(i)}, \ldots, x_{n_i}^{(i)}\}. \tag{4.57}$$

The truth value or extent of satisfaction of a proposition $x_j^{(i)} \varepsilon X^{(i)}$, is given by

$$h(x_j^{(i)}) \ \varepsilon\ [0,1]. \tag{4.58}$$

Additionally, the importances of groups of propositions in the premise, $X^{(i)}$, must be provided. They follow the constraints imposed on monotonic measures (see Definition 4.2.1):

$$g^{(i)} \ \varepsilon\ [0,1] \tag{4.59a}$$

$$g^{(i)}(\emptyset) = 0 \tag{4.59b}$$

$$g^{(i)}(X^{(i)}) = 1 \tag{4.59c}$$

and

$$\text{for } Q_1^{(i)}, Q_2^{(i)} \in P(X^{(i)}), \quad \text{if } Q_1^{(i)} \subseteq Q_2^{(i)},$$

$$g^{(i)}(Q_1^{(i)}) \leq g^{(i)}(Q_2^{(i)}). \tag{4.59d}$$

There are 2^{n_i} different arguments for $g^{(i)}$, and there are 2^{n_i} corresponding values.

The premise evaluation, $E_s(X^{(i)})$, is given by

$$E_s(X^{(i)}) = \oint_{X^{(i)}} h(x^{(i)}) \circ g^{(i)}(.)$$

$$= \operatorname*{Max}_{F \in P(X^{(i)})} \left(\operatorname*{Min}_{x_j^{(i)} \in F} (h(x_j^{(i)})) \wedge g^{(i)}(X^{(i)} \cap F) \right)$$

$$= \operatorname*{Max}_{F \in P(X^{(i)})} \left(\operatorname*{Min}_{x_j^{(i)} \in F} (h(x_j^{(i)})) \wedge g^{(i)}(F) \right). \tag{4.60}$$

For situations in which it is not possible to provide truth values to all the propositions in a premise, $X^{(i)}$, the integration is performed over a subset $Q^{(i)}$ of $X^{(i)}$ [The set $Q^{(i)}$ contains all $x_j^{(i)}$ for which truth values are available.]. Thus,

$$h(x_j^{(i)}) = 0, \quad \text{if } x_j^{(i)} \notin Q^{(i)}, \tag{4.61}$$

and the evaluation of the incomplete premise, $Q^{(i)}$, is given by

$$E_s(Q^{(i)}) = \oint_{Q^{(i)}} h(x^{(i)}) \circ g^{(i)}(.)$$

$$= \max_{F \varepsilon P(X^{(i)})} \left(\min_{x_j^{(i)} \varepsilon F} (h(x_j^{(i)})) \wedge g^{(i)}(Q^{(i)} \cap F) \right), \tag{4.62}$$

where, in general,

$$E_s(Q^{(i)}) \leq E_s(X^{(i)}). \tag{4.63}$$

The certainties or truth values for the actions, $Y_1^{(i)}, Y_2^{(i)}, \ldots, Y_{m_i}^{(i)}$, are given by

$$h(Y_k^{(i)}) = \Psi^{(i)}(E_s(Q^{(i)})), \quad k = 1, 2, \ldots, m_i , \tag{4.64}$$

where $\Psi_k^{(i)}$ is the certainty function for the k-th action in the i-th rule, and $E_s(Q^{(i)})$, $Q^{(i)} \subseteq X^{(i)}$, is the premise evaluation.

The rule base of the proposed production system would contain several production rules of the generic form presented above. During the operation of this production system, the interpreter schedules the execution of the rules. The sequence of rule selection, validation, and firing essentially nests the rules, and this elicits a chain of reasoning. The execution of each rule represents a single deductive step. But when rules are nested, the single deductive steps are chained, and the reasoning process spans several deductive steps.

The present formulation employs the Sugeno Integral to evaluate the premises of production rules. This functional permits the use of meta-level information in the evaluation, and is a generalization of the minimum operator. It can be used in conjunction with conventional data-driven and goal-driven strategies. The following

example demonstrates the application of the Sugeno Integral in multilevel reasoning.

Example 4.3.1

In the preceding section, we have seen that knowledge about a running back is specified by the production;

Rule 1: IF a man is well-built, but not too bulky ($x_1^{(1)}$)
AND he is very quick ($x_2^{(1)}$)
AND he has excellent ball-handling ability ($x_3^{(1)}$)
THEN he would make a good running back ($Y_1^{(1)}$)

or

Rule 1: IF $X^{(1)}$ [$g^{(1)}(.)$]
THEN $Y_1^{(1)}$ [$\Psi_1^{(1)}$]

where

$$X^{(1)} = \{x_1^{(1)}, x_2^{(1)}, x_3^{(1)}\}.$$

The measures of importance, $g^{(1)}(.)$, are as follows;

$$g^{(1)}(\emptyset) = 0$$
$$g^{(1)}(\{x_1^{(1)}\}) = 0.10$$
$$g^{(1)}(\{x_2^{(1)}\}) = 0.25$$
$$g^{(1)}(\{x_3^{(1)}\}) = 0.20$$
$$g^{(1)}(\{x_1^{(1)}, x_2^{(1)}\}) = 0.50$$
$$g^{(1)}(\{x_1^{(1)}, x_3^{(1)}\}) = 0.40$$
$$g^{(1)}(\{x_2^{(1)}, x_3^{(1)}\}) = 0.60$$

and

$$g^{(1)}(\{x_1^{(1)}, x_2^{(1)}, x_3^{(1)}\}) = 1.$$

The certainty function, $\Psi_1^{(1)}$, is *uniformly-true*, or *u-true* (see, e.g., Zadeh, 1975c), and is given by

$$\Psi_1^{(1)} = \Psi_1^{(1)}(\, E_s(Q^{(1)})\,)$$

$$= E_s(Q^{(1)}),$$

where

$$Q^{(1)} \subseteq X^{(1)},$$

and

$$E_s(Q^{(1)}) = \oint_{Q^{(1)}} h(x^{(1)}) \circ g^{(1)}(.).$$

Let us assume that there exists another rule that defines the attribute of build. This is given by

Rule 2: IF a man has the requisite height ($x_1^{(2)}$)
AND he has the requisite weight ($x_2^{(2)}$)
THEN he is well-built, but not too bulky ($Y_1^{(2)}$)

or

Rule 2: IF $X^{(2)}$ $[\ g^{(2)}(.)\ ]$
THEN $Y_1^{(2)}$ $[\ \Psi_1^{(2)}\ ]$

where

$$X^{(2)} = \{x_1^{(2)},\ x_2^{(2)}\}.$$

The measures of importance, $g^{(2)}(.)$, are as follows;

$$g^{(2)}(\emptyset) = 0$$

$$g^{(2)}(\{x_1^{(2)}\}) = 0.30$$

$$g^{(2)}(\{x_2^{(2)}\}) = 0.40$$

and

$$g^{(2)}(\{x_1^{(2)},\ x_2^{(2)}\}) = 1.$$

The certainty function, $\Psi_1^{(2)}$, is u-true, and is given by

$$\Psi_1^{(2)} = \Psi_1^{(2)}(\ E_s(Q^{(2)})\ )$$

$$= E_s(Q^{(2)}),$$

where

$$Q^{(2)} \subseteq X^{(2)},$$

and

$$E_s(Q^{(2)}) = \oint_{Q^{(2)}} h(x^{(2)}) \circ g^{(2)}(.).$$

Additionally, assume that the following rule defines the speed criterion;

Rule 3: IF a man has a good timing for the 40 yd. dash ($x_1^{(3)}$)
AND he has a good timing for the shuttle run ($x_2^{(3)}$)
AND he has a good timing for the obstacle race ($x_3^{(3)}$)
THEN he is very quick ($Y_1^{(3)}$)

or

Rule 3: IF $X^{(3)}$ [$g^{(3)}(.)$]
THEN $Y_1^{(3)}$ [$\Psi_1^{(3)}$]

where

$$X^{(3)} = \{x_1^{(3)}, x_2^{(3)}, x_3^{(3)}\}.$$

The measures of importance, $g^{(3)}(.)$, are as follows;

$$g^{(3)}(\emptyset) = 0$$

$$g^{(3)}(\{x_1^{(3)}\}) = 0.20$$

$$g^{(3)}(\{x_2^{(3)}\}) = 0.20$$

$$g^{(3)}(\{x_3^{(3)}\}) = 0.15$$

$$g^{(3)}(\{x_1^{(3)}, x_2^{(3)}\}) = 0.60$$
$$g^{(3)}(\{x_1^{(3)}, x_3^{(3)}\}) = 0.50$$
$$g^{(3)}(\{x_2^{(3)}, x_3^{(3)}\}) = 0.50$$

and

$$g^{(3)}(\{x_1^{(3)}, x_2^{(3)}, x_3^{(3)}\}) = 1.$$

The certainty function, $\Psi_1^{(3)}$, is u-true, and is given by

$$\Psi_1^{(3)} = \Psi_1^{(3)}(\ E_s(Q^{(3)})\ )$$
$$= E_s(Q^{(3)}),$$

where

$$Q^{(3)} \subseteq X^{(3)},$$

and

$$E_s(Q^{(3)}) = \oint_{Q^{(3)}} h(x^{(3)}) \circ g^{(3)}(.).$$

Note that the actions of Rules 2 and 3 are propositions in the premise of Rule 1. The three rules are linked, and the tree that is generated is presented in Figure 4.1.

Let us suppose that a walk-on candidate has been evaluated as satisfying the propositions to the following levels;

$h(x_3^{(1)}) = 0.35$ (ball-handling ability)

$h(x_1^{(2)}) = 0.45$ (height)

$h(x_2^{(2)}) = 0.55$ (weight)

$h(x_1^{(3)}) = 0.60$ (40 yard dash timing)

$h(x_2^{(3)}) = 0.55$ (shuttle course timing)

and

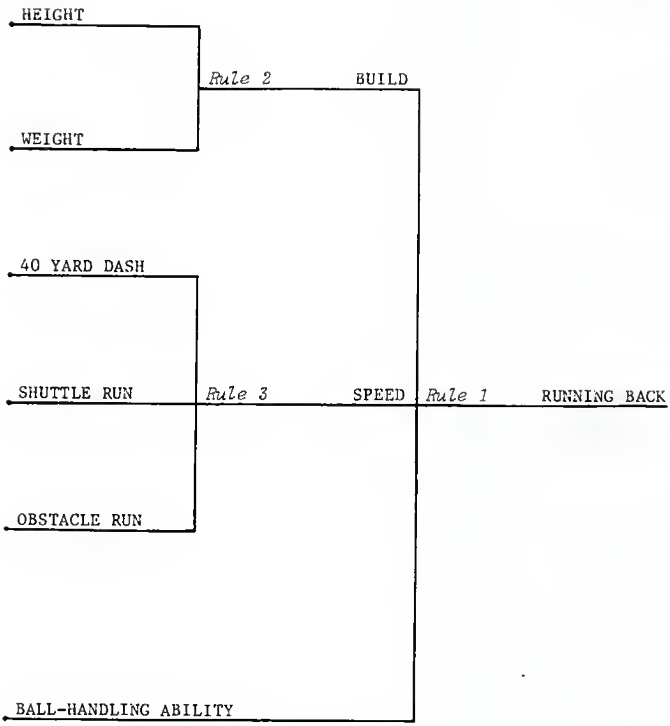


Figure 4.1. Selection of a running back.

$$h(x_3^{(3)}) = 0.75 \text{ (obstacle course timing).}$$

The process begins by evaluating Rule 2. Since truth values have been provided for all the propositions in the premise, integration is performed over the premise set, $X^{(2)}$ [i.e., $Q^{(2)} = X^{(2)}$]. The premise evaluation is given by

$$\begin{aligned} E_s(X^{(2)}) &= \oint_{X^{(2)}} h(x^{(2)}) \circ g^{(2)}(.) \\ &= 0.45. \end{aligned}$$

The truth value or certainty of the corresponding action, $Y_1^{(2)}$, is given by

$$\begin{aligned} h(Y_1^{(2)}) &= \Psi_1^{(2)}(E_s(X^{(2)})) \\ &= E_s(X^{(2)}) \\ &= 0.45. \end{aligned}$$

The certainty of the action, $Y_1^{(2)}$, is also the degree to which the candidate satisfies the build criterion in Rule 1, i.e.,

$$\begin{aligned} h(x_1^{(1)}) &= h(Y_1^{(2)}) \\ &= 0.45. \end{aligned}$$

Next, Rule 3 is evaluated. Since truth values have been provided for all the propositions in its premise, integration is performed over the premise set, $X^{(3)}$ [i.e., $Q^{(3)} = X^{(3)}$]. The evaluation is given by

$$\begin{aligned} E_s(X^{(3)}) &= \oint_{X^{(3)}} h(x^{(3)}) \circ g^{(3)}(.) \\ &= 0.55. \end{aligned}$$

The truth value or certainty of the corresponding action, $Y_1^{(3)}$, is

given by

$$h(Y_1^{(3)}) = \Psi_1^{(3)}(\ E_s(X^{(3)})\ ) = E_s(X^{(3)}) = 0.55.$$

This certainty is also the degree to which the candidate satisfies the speed criterion in Rule 1, i.e.,

$$h(x_2^{(1)}) = h(Y_1^{(3)}) = 0.55.$$

Finally, Rule 1 is evaluated. Since truth values for its propositions are available, integration is performed over the premise set, $X^{(1)}$ [i.e., $Q^{(1)} = X^{(1)}$]. The evaluation is given by

$$E_s(X^{(1)}) = \oint_{X^{(1)}} h(x^{(1)}) \circ g^{(1)}(.) = 0.45.$$

The truth value for the action, $Y_1^{(1)}$, is given by

$$h(Y_1^{(1)}) = \Psi_1^{(1)}(\ E_s(X^{(1)})\ ) = E_s(X^{(1)}) = 0.45,$$

or, the walk-on candidate has been found to satisfy our conception of a good running back to a level of 0.45.

Note that the levels to which the candidate has satisfied the propositions; $x_3^{(1)}$, $x_1^{(2)}$, $x_2^{(2)}$, $x_1^{(3)}$, $x_2^{(3)}$, and $x_3^{(3)}$, have been provided. There are no missing data, and all evaluations are complete. In a real-world problem, there is always a possibility that some data

may be missing. In such instances, we suggest that default truth values of zero be assigned to propositions for which data are not available. Since integrations are performed over subsets of premise sets, one or more premise evaluations are essentially incomplete; and the final evaluation (the extent to which a candidate satisfies our conception of a good running back) is even less certain.

CONCLUDING REMARKS

The production rule formalism is natural to human strategies of problem solving and decision making in several areas of human activity. This has contributed to its increased use in conventional expert systems.

Human beings are able to engage in approximate reasoning. In fact, it is this ability to reason in imprecise, subjective terms that distinguishes human intelligence from machine intelligence. The specification of premises of production rules in terms of soft propositions that can take *shades* of truth, represents an attempt to induce human-like reasoning that is able to cope with the imprecision rampant in the real-world.

The goal of research in AI is to emulate intelligent human activity. Yet, it is important to realize that this does not mean that a computer program is considered successful only if it duplicates human intelligence in its entirety. Instead, the intention must be to simulate human strategies as reasonably as possible. Given the subjectivity that enters into human decision making, it is evident that the specification of soft production rules, on its own, is not sufficient. Human beings tend to weigh and balance the individual pieces of evidence in arriving at an evaluation. Clearly, a functional is needed to model this subjective combination of evidence. The evaluation of premises of production rules using the Sugeno Integral is, therefore, an attempt to introduce the human quality of subjective combination of evidence into the production

rule formalism.

In our methodology, premises are written in the form of sets of AND-connected propositions. This is done without any loss of generality. Monotonic measures of sets represent the magnitudes of importance that groups of propositions (or subsets of premises) carry. These measures are meta-level descriptions of the *a priori* notions that are inherent in human expertise. Thus, in our attempt to introduce human subjectivity into mechanistic decision making, the premise evaluation has been decomposed into two distinct parts. One is the intrinsic importance that propositions carry in an evaluation, and the other, is the extent to which the propositions are satisfied. The Sugeno Integral combines these two aspects non-linearly. The result is a *mean* evaluation that has excellent intuitive justification. Additionally, the Sugeno Integral, which includes the conventional minimum operator as a special case, also provides a convenient framework for modeling the conservatism that is seen in human reasoning in the presence of ignorance.

The knowledge-based strategy in AI is a pragmatic approach towards the emulation of intelligent human activity. The pragmatism exhorts us to look for better ways to express, recognize, and use diverse and particular forms of knowledge in solving realistic problems that have been suitably constrained so that useful solutions are obtained. We believe that the present methodology is in line with this pragmatism.

REFERENCES

1. Barr, A., and Feigenbaum, E. A., "The Handbook of Artificial Intelligence: Vol. 1", William Kaufman Inc., Los Altos, Calif. (1981).

2. Barr, A., and Feigenbaum, E. A., "The Handbook of Artificial Intelligence: Vol. 2", William Kaufman Inc., Los Altos, Calif. (1982).

3. Cohen, P. R., and Feigenbaum, E. A., "The Handbook of Artificial Intelligence: Vol. 3", William Kaufman Inc., Los Altos, Calif. (1982).

4. Dageforde, G. L., Personal communication (1983).

5. Davis, R., and King, J. J., "An Overview of Production Systems", In Elcock, E., and Michie, D., (Eds.), "Machine Intelligence 8", Ellis Horwood, Chichester, U.K., 300-332 (1977).

6. Duda, R. O., and Shortliffe, E. H., "Expert Systems Research", Science, 220, No. 4594, 261-268 (1983).

7. Dybing, K. D., Personal communication (1983).

8. Gevarter, W. B., "Expert Systems: limited but powerful", IEEE Spectrum, 20, No. 8, 39-45 (1983).

9. McDermott, J., "Rl, a rule-based configurer of computer systems", Artif. Intell., 19, 38-88 (1982).

10. Newell, A., and Simon, H. A., "Human Problem Solving", Prentice-Hall, Engleswood Cliffs, N. J. (1972).

11. Nilsson, N. J., "Principles of Artificial Intelligence", Tioga Publishing Company, Palo Alto, Calif. (1980).

12. Post, E. A., "Formal reductions of the general combinatorial problem", Amer. J. Math., 65, 197-215 (1943).

13. Shafer, G., "A Mathematical Theory of Evidence", Princeton Univ. Press, Princeton, N. J. (1976).

14. Shortliffe, E. H., "Computer-based Medical Computations: MYCIN", American Elsevier, New York (1976).

15. Sugeno, M., "Theory of Fuzzy Integrals and its Applications", Ph.D. Thesis, Tokyo Institute of Technology, Tokyo (1974).

16. Zadeh, L. A., "The concept of a linguistic variable and its application to approximate reasoning, Part I", Inf. Sc., 8, 199-249 (1975a).

17. Zadeh, L. A., "The concept of a linguistic variable and its application to approximate reasoning, Part II", Inf. Sc., 8, 301-357 (1975b).

18. Zadeh, L. A., "The concept of a linguistic variable and its application to approximate reasoning, Part III", Inf. Sc., 9, 43-80 (1975c).

CHAPTER V

AN APPLICATION IN THE CLASSIFICATION OF RICE GRAIN

Processing operations performed on rice crops give rise to appreciable amounts of broken grain. Unlike other cereals, rice demands a premium price only as whole grain. Broken fragments sell for approximately half the price of corresponding whole kernels, and in the rice industry, it is a matter of considerable importance to divide rice grain into several classes, depending on the degree to which the kernels are damaged. The emphasis on uniformity and accuracy in specifying standards on rice quality has seen the acceptance of mechanical procedures of classification that involve the use of sieves and plates. Although these procedures do work rather well, there is no doubt that they tend to move away from a common-sense view of rice classification.

The task of separating rice grains into different classes depending on the extent of breakage, we believe, is purely a matter of visual discrimination that is easily and effectively performed by trained personnel. Observation of experienced grain inspectors indicates that they have clear notions of what prototypes from different classes look like. More specifically, they know what to look for, and exactly how to assess what they see. In performing their analyses, they draw and act on relevant pieces of domain-specific knowledge.

In this chapter, the ideas developed in preceding chapters are brought to bear on the problem of classification of rice grain. We attempt to follow the visual approach that an expert grain inspector would adopt. Information specific to the task of classification is coded in the form of production rules (see, e.g., Barr and Feigenbaum,

1981). Essentially, a prototype from each class is defined by a unique production rule, similar in form to a *discriminant function* (see, e.g., Andrews, 1972) employed in classical pattern recognition theory.

ON THE CLASSIFICATION OF RICE GRAIN

The quality of rice is usually evaluated on the basis of its suitability for a specific end-use by a particular group of consumers. In the United States, where standards on quality are rigidly observed, rice quality depends on a variety of parameters, such as grain size, shape, uniformity and general appearance. Other pertinent factors concern milling yields, cooking and processing characteristics, cleanliness, soundness and purity (Webb and Stermer, 1972). Throughout much of the rice-growing world, however, the standards on quality are less stringent, and there is still no generally accepted basis for evaluation. Yet, since most rice is processed and consumed in whole kernel form, we may anticipate that the physical properties of the intact kernel, such as shape, size, and general appearance, are of particular significance in determining the quality of rice. Indeed, it must be emphasized that the majority of international trade in the commodity is generally conducted on the basis of quality as determined by visual examination by experts (Webb and Stermer, 1972).

Rice is unique among cereals in that it is almost always used for human consumption, and demands a premium price only as whole grain. A certain amount of breakage, however, is unavoidable. Rice kernels have been observed to be cracked while still in the husk, and are also broken during harvesting, handling, drying, and milling. Milling, which is almost always performed on rough rice (unhulled rice, also known as *paddy*), involves the removal of the hull, bran layers and germ, while

preserving the kernels to approximately their original shape. It is of interest to note that about 15 percent of all rice milled in the United States is broken (U.S.D.A., 1969). This translates to an annual figure of approximately 1 billion lbs. [0.45 million metric tons]. Since the larger pieces of broken rice sell for a little more than one-half the price of whole rice, while smaller fragments sell for less than one-half the price of comparable whole grains; the annual loss due to breakage may be conservatively estimated to be around 30 million dollars. A great deal of effort has already been focussed toward the reduction of grain breakage, and many experts opine that the percentage of grain breakage will not reduce drastically. It is, therefore, quite obvious that economic considerations necessitate the adoption of reliable and efficient procedures for separating whole kernels from broken fragments of rice.

In the United States, rice is marketed under three types, designated as *long-grain*, *medium-grain*, and *short-grain* (Adair, 1972). The present work is confined in its scope to the treatment of *long-grain milled rice* (see APPENDIX I, for U.S. Standards for Milled Rice.).

Milling operations result in the production of rice grains with varying degrees of breakage. Depending on the extent of breakage, it is customary to divide milled rice into four classes; *wholes* or *head rice*, *second heads*, *screenings*, and *brewers* (see APPENDIX I). The class, wholes, consists largely of unbroken kernels, although, it is common to consider grains upto three-quarters of unbroken kernels as belonging to this category. Second heads are the larger broken grains, one-half to three-

quarters of the whole kernel. Screenings consists of medium-sized broken kernels, one-quarter to one-half of the whole grain. The smaller broken fragments that do not meet the kernel size requirements of the other classes are termed as brewers. It is conventional to designate fragments approximately one-quarter (or slightly lower) of the whole kernel as belonging to this category. Second heads and screenings are usually blended back into the head rice and used for human consumption (Witte, 1972). The extent of blending depends considerably on the orders placed by specific consumers, and the standards governing the quality of rice sold in the United States. Brewers, the least expensive of the four classes, is usually sold to breweries where it is used as an adjunct in the manufacture of beer (In the United States, it is permitted to use rice as a starch source for making beer.). Representative samples belonging to the four classes are presented in Figure 5.1.

Cursory examination of the grain samples seems to indicate that properties such as shape, size, and general appearance are sufficient for the task of distinguishing between grains belonging to different classes. More importantly, the task appears to be purely a problem of visual discrimination that could be performed by trained humans. However, the U.S. Standards for Milled Rice suggest the use of procedures that involve the operation of sieves and plates for classification. This is largely due to a need for uniformly accurate procedures for judging rice quality. In rice trading circles, there is a common feeling that unless the grading procedures are objective, the criteria of quality cannot be accurately and uniformly measured, interpreted, and specified.

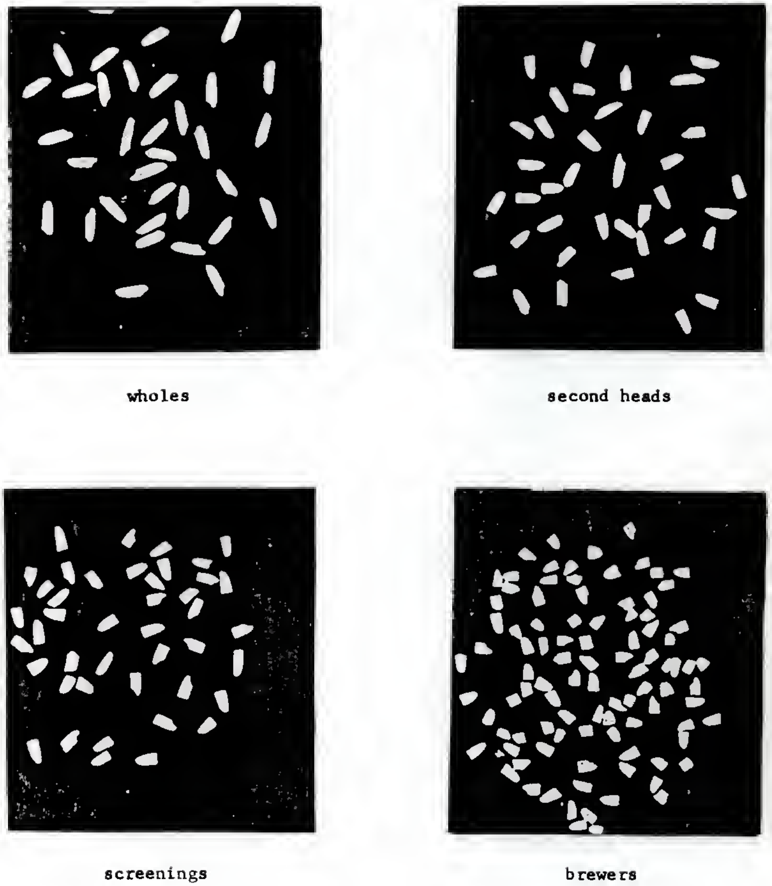


Figure 5.1. Representative samples from the four classes of rice grain.

It appears that the emphasis on objective procedures arises out of a lack of faith in the reproducibility of human judgment. We do not feel that this lack of faith is sufficient to justify the transmutation of the classification problem, from one viewed in common-sense visual terms, to another defined on the basis of objective and procedural aspects.

We must retain our *a priori* conceptions of grain classification. Prototypes from the four classes should be defined in terms of what we see, and not how we could mechanically separate them. In order to do so, we must attempt to understand and emulate the reasoning an expert grain inspector would adopt when he places a rice kernel on his palm, prior to making a quantitative assessment of quality.

PRODUCTION RULES FOR THE CLASSIFICATION OF RICE GRAIN

An experienced grain inspector, on examination of a rice kernel that rests on his palm, is almost effortlessly able to place it in one of the four prescribed classes. The assessment of the general appearance of a grain sample plays a major role in the process of discrimination. Production rules offer a convenient representation of knowledge relevant to the discrimination process. In this section, we attempt to construct the production rules that define prototypes from the four classes.

The extent of damage to the rice kernel is a basic attribute that determines the belonging of a grain sample to a given class. This enables us to specify the following simple rules.

Rule 1. IF the kernel is *approximately intact*
THEN the sample is a *whole*.

Rule 2. IF the kernel is *approximately three-quarters intact*
THEN the sample is a *second head*.

Rule 3. IF the kernel is *approximately one-half intact*
THEN the sample is a *screening*.

Rule 4. IF the kernel is *approximately one-quarter intact*
THEN the sample is a *brewer*.

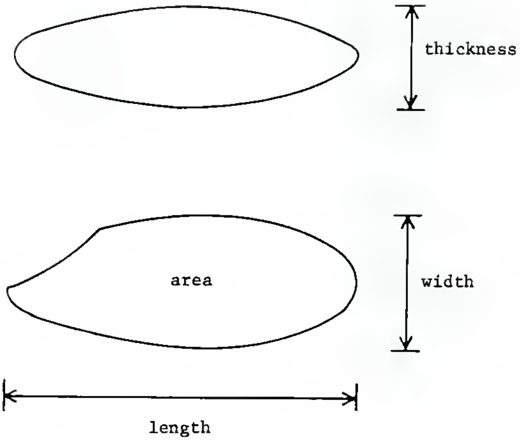
Considerable clarifications are needed before the general rules presented above can be applied. Specifically, it is important to understand precisely what the concepts *approximately intact*, *approximately three-quarters intact*, etc., entail. We contend that an expert grain inspector has the ability to solve the classification problem because he has clear notions of these concepts. Our task, therefore, is to elicit

this information from him and other experts.

There appear to be two separate strategies that a grain inspector uses when he attempts to evaluate the degree of *intactness*. In one scheme, he superimposes a mental image of a prototype kernel over the sample under consideration. Alternatively, he might extrapolate the sample to achieve the prototype grain, and then assess the extent of breakage. Consultations with several grain inspectors and grain processing experts have indicated that both schemes are employed, simultaneously, and to varying levels. The simulation of hybrid matching-extrapolation schemes requires sophisticated hardware and software capabilities. In our attempt to simplify the problem, while retaining these aspects of the discrimination process, we examine a set of more basic attributes, the consideration of which is implicit in both methodologies.

Each of the parameters, length, width, area, and slenderness, has an important bearing in an expert's assessment of the general appearance of a rice grain. Length is the distance between the most distant tips of a kernel. For milled rice, width is defined as the distance across the kernel at its widest point. The area that a kernel projects along a view, perpendicular to the length and width axes, is the area parameter. Slenderness is defined as the ratio of kernel length to its width. The four parameters, illustrated schematically in Figure 5.2, form the basis for definitions of the ambiguous concepts, *approximately intact*, etc., which occur in the simple discrimination rules specified above.

The incorporation of the four parameters, described in the preceding



$$\text{slenderness} = \frac{\text{length}}{\text{width}}$$

Figure 5.2. Shape parameters for milled rice kernels.

paragraph, in a model of visual discrimination is justifiable. Whether an expert superimposes a mental image of a prototype kernel over the grain sample, or, whether he extrapolates from the sample kernel, he proceeds from *a priori* concepts. In the first case, the mental image itself is the prior notion, and this could be specified as a function of the four parameters. On the other hand, we could consider the four parameters as governing the extent of extrapolation in the alternate scheme, in much the same way as strings control the movements of a marionette. Our simplified four parameter model, like a crude puppet, is unable to simulate human ability completely, and it may be necessary to introduce additional parameters at a later stage.

Notions about degrees of intactness, that appear in the simple rules of discrimination, may now be expressed in terms of the four parameters. The definitions, constructed in collaboration with experts, incorporate linguistic concepts to qualify the parameters of length, width, area, and slenderness (The grain experts who were consulted, found it convenient to express their knowledge in qualitative, linguistic terms.). These definitions are quite obvious, and are natural to human understanding of the discrimination problem.

approximately intact (wholes) implies
i) a high length parameter
and ii) a medium to high width parameter
and iii) a high area parameter
and iv) a high slenderness parameter.

approximately three-quarters intact (second heads) implies
i) a medium to high length parameter
and ii) a medium to high width parameter
and iii) a medium to high area parameter
and iv) a medium to high slenderness parameter.

approximately one-half intact (screenings) implies
i) a low to medium length parameter
and ii) a medium to high width parameter
and iii) a medium area parameter
and iv) a low to medium slenderness parameter.

approximately one-quarter intact (brewers) implies
i) a low length parameter
and ii) a medium to high width parameter
and iii) a low area parameter
and iv) a low slenderness parameter.

Corresponding to the four definitions, a set of four discrimination rules may be specified. They are given below.

Rule 1. IF the kernel has a high length parameter
AND the kernel has a medium to high width parameter
AND the kernel has a high area parameter
AND the kernel has a high slenderness parameter
THEN the kernel is a *whole*.

Rule 2. IF the kernel has a medium to high length parameter
AND the kernel has a medium to high width parameter
AND the kernel has a medium to high area parameter
AND the kernel has a medium to high slenderness parameter
THEN the kernel is a *second head*.

Rule 3. IF the kernel has a low to medium length parameter
AND the kernel has a medium to high width parameter
AND the kernel has a medium area parameter
AND the kernel has a low to medium slenderness parameter
THEN the kernel is a *screening*.

Rule 4. IF the kernel has a low length parameter
AND the kernel has a medium to high width parameter
AND the kernel has a low area parameter
AND the kernel has a low slenderness parameter
THEN the kernel is a *brewer*.

Symbolically, a rule may be written in the form:

$$\text{IF} \quad X^{(i)} \quad \text{THEN} \quad Y^{(i)}, \quad i = 1, 2, 3, 4 \tag{5.1a}$$

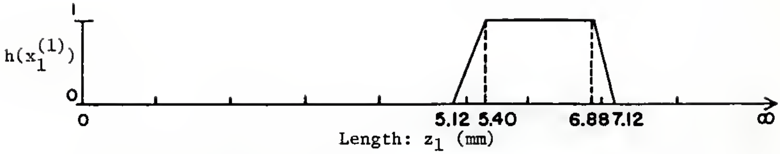
where

$$X^{(i)} = \{x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, x_4^{(i)}\} \tag{5.1b}$$

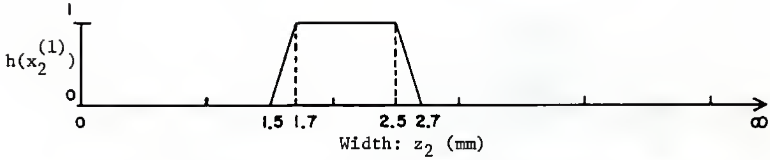
is the premise set consisting of four AND-connected propositions (pertaining to the length, width, area, and slenderness parameters of the kernel), and $Y^{(i)}$ is the action corresponding to the i-th discrimination rule.

Each rule defines a prototype kernel from one of the prescribed classes. Although, the rules do provide reasonable conceptions of the prototypes, they are still fairly ambiguous. The propositions that comprise the premises of the discrimination rules are subject to interpretation of the linguistic terms they contain. Before the rules can be applied, it is important that we specify exactly what the propositions entail.

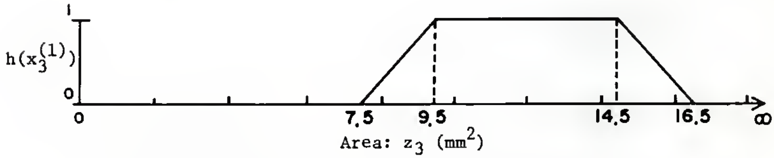
A clear understanding of the propositions, therefore, is an essential part of the expertise involved in grain classification. This understanding enables a grain inspector to evaluate correctly the extent to which a sample kernel satisfies each proposition, or more precisely, assign a satisfactory truth value to each proposition. In the present work, *trapezoidal representations* are employed to specify truth values, $h(x_j^{(i)})$, corresponding to a proposition $x_j^{(i)}$ in the i-th rule. The representations, illustrated in Figures 5.3 through 5.6, prescribe truth values (ranging from "0", or *false*, to "1", or *true*) over feasible ranges of the measured parameters. The truth value representations have been obtained in consultation with experts. The task involved the measurement of length, width, area, and slenderness for a large number of pre-evaluated kernels from each class (known as



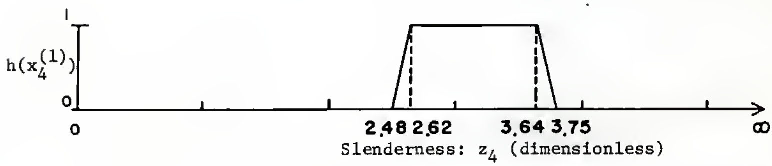
Proposition: *the kernel has a high length parameter*



Proposition: *the kernel has a medium to high width parameter*

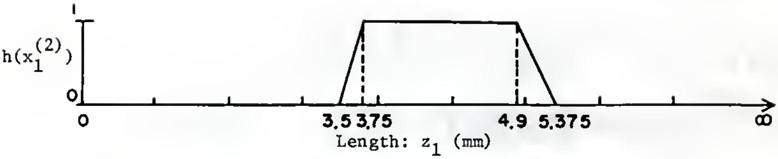


Proposition: *the kernel has a high area parameter*

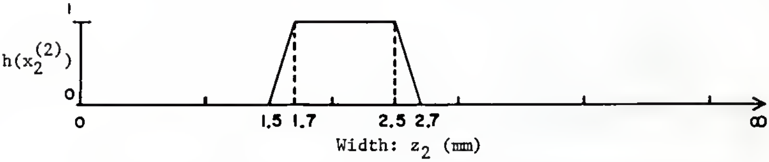


Proposition: *the kernel has a high slenderness parameter*

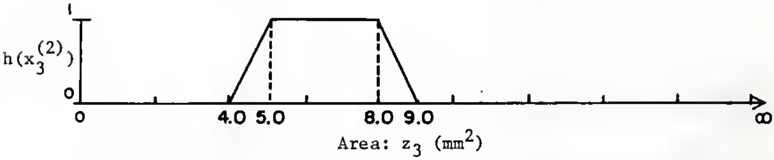
Figure 5.3. Truth values for the propositions defining a prototype whole.



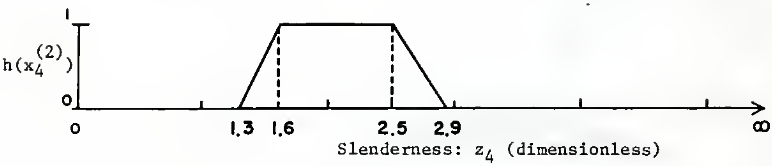
Proposition: *the kernel has a medium to high length parameter*



Proposition: *the kernel has a medium to high width parameter*

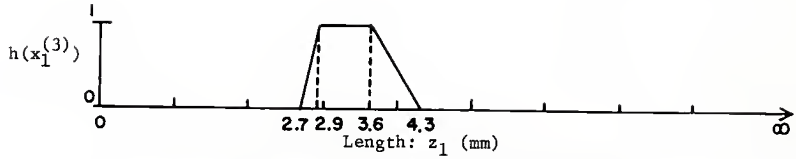


Proposition: *the kernel has a medium to high area parameter*

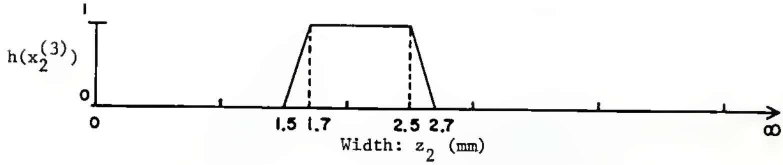


Proposition: *the kernel has a medium to high slenderness parameter*

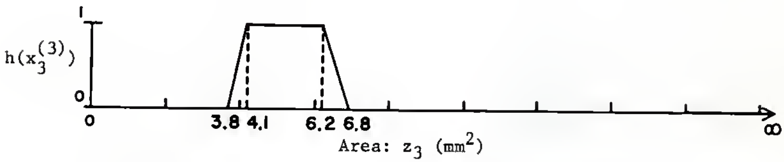
Figure 5.4. Truth values for the propositions defining a prototype second head.



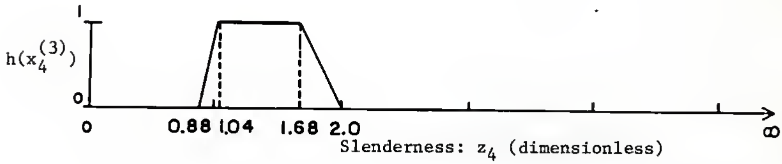
Proposition: *the kernel has a low to medium length parameter*



Proposition: *the kernel has a medium to high width parameter*

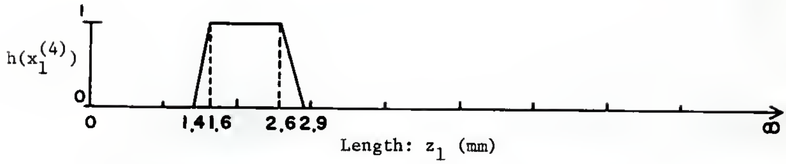


Proposition: *the kernel has a medium area parameter*

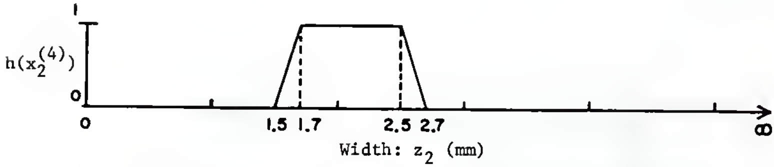


Proposition: *the kernel has a low to medium slenderness parameter*

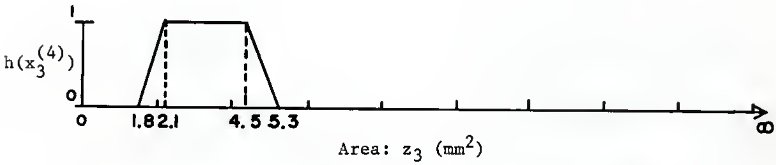
Figure 5.5. Truth values for the propositions defining a prototype screening.



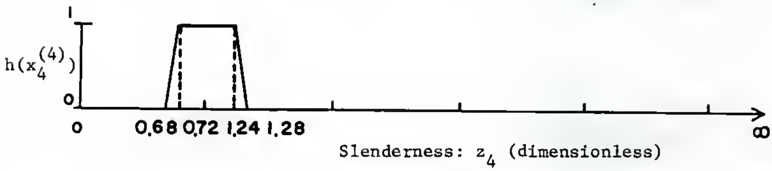
Proposition: *the kernel has a low length parameter*



Proposition: *the kernel has a medium to high width parameter*



Proposition: *the kernel has a low area parameter*



Proposition: *the kernel has a low slenderness parameter*

Figure 5.6. Truth values for the propositions defining a prototype brewer.

a *training set*, in the jargon of pattern recognition). Using this information, a range of variation for the measured parameter corresponding to each proposition was determined. The guidance of experts was solicited for the purpose of specifying truth values over each range.

It is now possible to assess the degrees to which the parameters of an unclassified kernel satisfy individual propositions. The premise of each rule, however, consists of four separate propositions, and the assessment of these propositions produces four distinct pieces of evidence which must be combined in evaluating the level to which the unclassified kernel satisfies the entire premise. Since the premises are comprised of AND-connected propositions, it is conventional to employ the minimum operator to perform the combination (see, e.g., Shortliffe, 1976; and McDermott, 1982). Using this scheme, the premise evaluation for the i-th rule, $E_m(X^{(i)})$, is given by

$$E_m(X^{(i)}) = \min_{x_j^{(i)} \in X^{(i)}} (h(x_j^{(i)})), \tag{5.2}$$

where $h(x_j^{(i)})$ is the degree to which the sample kernel satisfies the j-th proposition in the i-th rule. The major drawback of this scheme is that it provides a very pessimistic evaluation. The sample kernel is judged to be only as satisfactory as its worst quality indicates, and it is debatable whether grain experts are so conservative in their evaluations.

Our conversations with grain experts have indicated that they

are inclined to weigh and balance the different pieces of evidence before arriving at an evaluation. A sample kernel may not satisfy one proposition very well; yet, if the other propositions are satisfied sufficiently, a trade-off is performed, and the resultant evaluation could be quite high. Information about the relative weights that propositions carry in an evaluation is significant in the trade-off procedure. Some propositions are more important than others, and a look at the discrimination rules will illustrate this point. Note that the proposition pertaining to the width parameter

the kernel has a medium to high width parameter

is common to all four rules. This is because only seldom in normal grain processing, does a rice kernel break along a plane parallel to the length axis (a break, so to speak, *against the grain*). Hence, although the width proposition may be pertinent for the task of differentiating rice kernels from other cereal grains, it does not provide any information for the specific task of classifying rice. On the other hand, the propositions dealing with length, area, and slenderness are of special significance in both tasks, and would surely carry greater weights in the evaluation of a premise.

The concept of monotonic measures, also known as fuzzy measures (Sugeno, 1974; also see the preceding chapters), is employed to capture the *a priori* notions that experts have concerning the relative weights of propositions. In the present methodology, the premise of the i-th rule, $X^{(i)}$, is written as a set, i.e.,

$$X^{(i)} = \{x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, x_4^{(i)}\}, \tag{5.3}$$

and the monotonic measure

$$g^{(i)}(Q^{(i)}) \in [0,1], \quad Q^{(i)} \subseteq X^{(i)}, \tag{5.4}$$

represents the combined importance that the set of propositions, $Q^{(i)}$, carries in an evaluation. The entire premise, $X^{(i)}$, corresponds to the totality of evidence, and

$$g^{(i)}(X^{(i)}) = 1. \tag{5.5}$$

On the other hand, the null set, $\emptyset$, contains no propositions, and, therefore, contributes no information to the evaluation. Thus,

$$g^{(i)}(\emptyset) = 0. \tag{5.6}$$

For other sets,

$$Q^{(i)} \subseteq X^{(i)}, \quad Q^{(i)} \neq X^{(i)}, \quad Q^{(i)} \neq \emptyset, \tag{5.7}$$

depending on the propositions contained; a value in the closed interval [0, 1] is assigned to the measure, $g^{(i)}(Q^{(i)})$. The assignment of values must follow the axiom of monotonicity, that is,

$$Q_1^{(i)}, Q_2^{(i)} \subseteq X^{(i)}, \quad Q_1^{(i)} \subseteq Q_2^{(i)},$$

$$g^{(i)}(Q_1^{(i)}) \leqslant g^{(i)}(Q_2^{(i)}). \tag{5.8}$$

The set, $Q^{(i)}$, is a subset of the premise, $X^{(i)}$, and for each rule, there are sixteen different subsets, or mathematically,

$$Q^{(i)} \in P(X^{(i)})^{*} \tag{5.9a}$$

---

* $P(X^{(i)})$ is known as the power set, or set of all subsets, of $X^{(i)}$. The cardinality (number of elements) of $P(X^{(i)})$, written as card(P($X^{(i)}$)), is 16.

where

$$P(X^{(i)}) = \{Q_1^{(i)}, Q_2^{(i)}, \ldots, Q_{16}^{(i)}\} \tag{5.9b}$$

$$Q_1^{(i)} = \emptyset \tag{5.9c}$$

$$Q_2^{(i)} = \{x_1^{(i)}\} \tag{5.9d}$$

$$Q_3^{(i)} = \{x_2^{(i)}\} \tag{5.9e}$$

$$Q_4^{(i)} = \{x_3^{(i)}\} \tag{5.9f}$$

$$Q_5^{(i)} = \{x_4^{(i)}\} \tag{5.9g}$$

$$Q_6^{(i)} = \{x_1^{(i)}, x_2^{(i)}\} \tag{5.9h}$$

$$Q_7^{(i)} = \{x_1^{(i)}, x_3^{(i)}\} \tag{5.9i}$$

$$Q_8^{(i)} = \{x_1^{(i)}, x_4^{(i)}\} \tag{5.9j}$$

$$Q_9^{(i)} = \{x_2^{(i)}, x_3^{(i)}\} \tag{5.9k}$$

$$Q_{10}^{(i)} = \{x_2^{(i)}, x_4^{(i)}\} \tag{5.9l}$$

$$Q_{11}^{(i)} = \{x_3^{(i)}, x_4^{(i)}\} \tag{5.9m}$$

$$Q_{12}^{(i)} = \{x_1^{(i)}, x_2^{(i)}, x_3^{(i)}\} \tag{5.9n}$$

$$Q_{13}^{(i)} = \{x_1^{(i)}, x_2^{(i)}, x_4^{(i)}\} \tag{5.9o}$$

$$Q_{14}^{(i)} = \{x_1^{(i)}, x_3^{(i)}, x_4^{(i)}\} \tag{5.9p}$$

$$Q_{15}^{(i)} = \{x_2^{(i)}, x_3^{(i)}, x_4^{(i)}\} \tag{5.9q}$$

and

$$Q_{16}^{(i)} = \{x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, x_4^{(i)}\} = X^{(i)}. \tag{5.9r}$$

The monotonic measure, $g^{(i)}(Q^{(i)})$, is used to represent the *a priori* notion of the weight carried by the body of evidence, $Q^{(i)}$, in an evaluation of the premise. Our task, therefore, is to assign values between zero and one, for the sixteen different bodies of evidence provided by each of the four rules (or, 16 x 4 = 64 values from the [0, 1] interval).

The experts we have consulted are of the opinion that since each rule deals with the same four parameters, it would be reasonable to assume that the measures are identical for equivalent sets from all rules. Thus,

$$g^{(1)}(Q_j^{(1)}) = g^{(2)}(Q_j^{(2)}) = g^{(3)}(Q_j^{(3)}) = g^{(4)}(Q_j^{(4)}),$$
$$j = 1, 2, \ldots, 16. \qquad (5.10)$$

Once this assumption is made, it is necessary to assign values to just sixteen measures; in reality, only fourteen values need be specified, since, by definition,

$$g^{(i)}(\emptyset) = 0, \qquad (5.6)$$

and

$$g^{(i)}(X^{(i)}) = 1. \qquad (5.5)$$

The experts have found it quite convenient to specify magnitudes of relative importance for individual propositions; however, the assignment of values to subsets containing more than one proposition is a difficult task. To alleviate this difficulty, the sixteen subsets are arranged in order of increasing importance. The values are then carefully assigned, one at a time, always checking that the axiom of monotonicity is satisfied. The following measures have been found to be suitable for the discrimination rules:

$$g^{(i)}(\emptyset) = 0 \qquad (5.11a)$$
$$g^{(i)}(\{x_2^{(i)}\}) = 0.01 \qquad (5.11b)$$
$$g^{(i)}(\{x_1^{(i)}\}) = 0.02 \qquad (5.11c)$$
$$g^{(i)}(\{x_4^{(i)}\}) = 0.03 \qquad (5.11d)$$
$$g^{(i)}(\{x_3^{(i)}\}) = 0.05 \qquad (5.11e)$$

$$g^{(i)}(\{x_1^{(i)}, x_2^{(i)}\}) = 0.30 \qquad (5.11f)$$

$$g^{(i)}(\{x_2^{(i)}, x_4^{(i)}\}) = 0.30 \qquad (5.11g)$$

$$g^{(i)}(\{x_1^{(i)}, x_4^{(i)}\}) = 0.30 \qquad (5.11h)$$

$$g^{(i)}(\{x_2^{(i)}, x_3^{(i)}\}) = 0.35 \qquad (5.11i)$$

$$g^{(i)}(\{x_1^{(i)}, x_3^{(i)}\}) = 0.40 \qquad (5.11j)$$

$$g^{(i)}(\{x_3^{(i)}, x_4^{(i)}\}) = 0.55 \qquad (5.11k)$$

$$g^{(i)}(\{x_1^{(i)}, x_2^{(i)}, x_4^{(i)}\}) = 0.40 \qquad (5.11\ell)$$

$$g^{(i)}(\{x_1^{(i)}, x_2^{(i)}, x_3^{(i)}\}) = 0.60 \qquad (5.11m)$$

$$g^{(i)}(\{x_2^{(i)}, x_3^{(i)}, x_4^{(i)}\}) = 0.65 \qquad (5.11n)$$

$$g^{(i)}(\{x_1^{(i)}, x_3^{(i)}, x_4^{(i)}\}) = 0.75 \qquad (5.11o)$$

and

$$g^{(i)}(\{x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, x_4^{(i)}\}) = 1. \qquad (5.11p)$$

The Sugeno Integral (Sugeno, 1974; also see the preceding chapters for additional details) combines the bodies of evidence provided by individual propositions. The resultant expression for the premise evaluation, $E_s(X^{(i)})$, is given by

$$E_s(X^{(i)}) = \oint_{X^{(i)}} h(x^{(i)}) \bullet g^{(i)}(.)$$

$$= \underset{F \in P(X^{(i)})}{\text{Max}} \left(\underset{x_j^{(i)} \in F}{\text{Min}} (h(x_j^{(i)})) \wedge g^{(i)}(X^{(i)} \cap F) \right)$$

$$= \underset{F \in P(X^{(i)})}{\text{Max}} \left(\underset{x_j^{(i)} \in F}{\text{Min}} (h(x_j^{(i)})) \wedge g^{(i)}(F) \right). \qquad (5.12)$$

The Sugeno Integral in Equation (5.12) is evaluated over the reference

set, $X^{(i)}$, and corresponds to an evaluation of the entire premise.

In certain situations, it may not be possible to evaluate a premise completely. Specifically, information pertaining to certain parameters may be missing, and truth values cannot be assigned to all propositions within a premise. The present methodology permits the partial evaluation of a premise. This may be represented by the evaluation of the sub-rule;

$$\text{IF} \quad Q^{(i)} \quad \text{THEN} \quad Y^{(i)} \tag{5.13a}$$

where

$$Q^{(i)} \subseteq X^{(i)}, \tag{5.13b}$$

for which

$$\begin{aligned} E_s(Q^{(i)}) &= \oint_{Q^{(i)}} h(x^{(i)}) \circ g^{(i)}(.) \\ &= \max_{F \varepsilon P(X^{(i)})} \left(\min_{x_j^{(i)} \varepsilon F} (h(x^{(i)})) \wedge g^{(i)}(Q^{(i)} \cap F) \right). \end{aligned} \tag{5.14}$$

The set, $Q^{(i)}$, contains the propositions for which data are available. We suggest that truth values of zero be assigned to propositions for which no information is available. This corresponds to the conservatism that is often seen in human analyses performed in the presence of ignorance. The conservatism also manifests itself in the resultant evaluation, $E_s(Q^{(i)})$. Since the Sugeno Integral is also monotonic, i.e.,

$$E_s(Q^{(i)}) \leq E_s(X^{(i)}), \quad \text{for } Q^{(i)} \subseteq X^{(i)}. \tag{5.15}$$

In other words, a partial evaluation, or one performed in a state of ignorance, is never better than the complete evaluation of the

premise.

To this point, we have confined ourselves to the evaluation of premises of the discrimination rules. A production rule, however, has two specific parts, a condition, and an action. Therefore, having obtained the premise evaluation, $E_s(Q^{(i)})$, it is necessary to translate this information into a certainty of the corresponding action. The certainty or truth value of an action,

$$h(Y^{(i)}) \in [0,1], \tag{5.16}$$

depends on the premise evaluation, and the strength of the link between the condition and action.

A certainty function, $\Psi^{(i)}$, is used to relate the action certainty to the evaluation of its premise (see Chapter IV). This function, which is similar to a certainty factor in MYCIN (Shortliffe, 1976), is adjusted to express the strength of the condition-action link. For the purposes of grain classification, we assume that the action of each rule is as certain as its premise evaluation. Thus, the certainty functions take a u-true, or uniformly-true form (see, e.g., Zadeh, 1975c); this gives rise to

$$\begin{aligned} h(Y^{(i)}) &= \Psi^{(i)}(E_s(Q^{(i)})) \\ &= E_s(Q^{(i)}), \quad i = 1, 2, 3, 4, \end{aligned} \tag{5.17a}$$

where

$$Q^{(i)} \subseteq X^{(i)}. \tag{5.17b}$$

A RULE-BASED CLASSIFIER FOR RICE GRAIN

In the preceding section, we have developed simple rules that define prototype kernels from the four prescribed classes of rice grain. The simple rules of discrimination are quite natural to an expert's understanding of the classification problem. Naturally, at this stage, it is of interest to implement these ideas in a simple rule-based classifier that relies on these rules to identify unknown kernels.

The following rules define the four classes;

Rule 1.	IF	$X^{(1)}$	THEN	$Y^{(1)}$	(*wholes*)
Rule 2.	IF	$X^{(2)}$	THEN	$Y^{(2)}$	(*second heads*)
Rule 3.	IF	$X^{(3)}$	THEN	$Y^{(3)}$	(*screenings*)
Rule 4.	IF	$X^{(4)}$	THEN	$Y^{(4)}$	(*brewers*)

(see the preceding section for details). Input data pertaining to the shape and size of an unclassified kernel, ℓ, is represented by a pattern vector, $\hat{z}_\ell$. This is given by

$$\hat{z}_\ell = (z_{\ell 1}, z_{\ell 2}, z_{\ell 3}, z_{\ell 4}), \qquad (5.18a)$$

where

$$z_{\ell 1} = \text{length of kernel } \ell \text{ (mm)} \qquad (5.18b)$$

$$z_{\ell 2} = \text{width of kernel } \ell \text{ (mm)} \qquad (5.18c)$$

$$z_{\ell 3} = \text{area of kernel } \ell \text{ (mm}^2\text{)} \qquad (5.18d)$$

$$z_{\ell 4} = \text{slenderness parameter of kernel } \ell \text{ (dimensionless)} \quad (z_{\ell 4} = z_{\ell 1}/z_{\ell 2}). \qquad (5.18e)$$

Values of zero are substituted in place of missing data in the

pattern vector.

The classifier proceeds by scheduling Rules 1 through 4 in sequence. For each rule i, depending on the elements in the pattern vector, $\hat{z}_\ell$, truth values,

$$h(x_j^{(i)})_\ell \in [0,1], \quad j = 1, 2, 3, 4, \tag{5.19}$$

are assigned to the propositions, $x_j^{(i)}$, in the premise. The Sugeno Integral is used to obtain the premise evaluation,

$$E_s(Q^{(i)}) \ , \quad Q^{(i)} \subseteq X^{(i)}; \tag{5.20}$$

after which, the corresponding action certainty, $h(Y^{(i)})_\ell$, is estimated. The certainty functions are all assumed to be uniformly-true, i.e.,

$$\begin{aligned} h(Y^{(i)})_\ell &= \Psi^{(i)}(E_s(Q^{(i)})_\ell) \\ &= E_s(Q^{(i)})_\ell, \end{aligned} \tag{5.21a}$$

where

$$Q^{(i)} \subseteq X^{(i)}. \tag{5.21b}$$

Rules 1 through 4 provide independent estimates of certainty for their actions, i.e.,

$h(Y^{(1)})_\ell \equiv$ certainty that kernel ℓ is a whole (Rule 1)

$h(Y^{(2)})_\ell \equiv$ certainty that kernel ℓ is a second head (Rule 2)

$h(Y^{(3)})_\ell \equiv$ certainty that kernel ℓ is a screening (Rule 3)

$h(Y^{(4)})_\ell \equiv$ certainty that kernel ℓ is a brewer (Rule 4).

The action certainty, $h(Y^{(i)})_\ell$, may be interpreted as representing the *degree of belonging* of the unclassified kernel, ℓ, to the class defined by the i-th rule. Therefore, we place the kernel, ℓ, in the class which has the highest certainty. A rule that prescribes the allocation of an unclassified kernel to a class must be scheduled

after the application of Rules 1 through 4. This rule is given by

Rule 5. IF $h(Y^{(k)})_{\ell} = \underset{i=1}{\overset{4}{\mathrm{Max}}} (h(Y^{(i)})_{\ell})$, $k \in \{1, 2, 3, 4\}$

THEN adopt the action $Y^{(k)}$
(or, kernel ℓ belongs to the class defined by Rule k).

In some instances, two or more action certainties could satisfy Rule 5. An additional rule is required to resolve deadlocks, i.e.,

Rule 6. IF $h(Y^{(k1)})_{\ell} = h(Y^{(k2)})_{\ell} = \underset{i=1}{\overset{4}{\mathrm{Max}}} (h(Y^{(i)})_{\ell})$,

AND $k1 < k2$, $k1, k2 \in \{1, 2, 3, 4\}$

THEN adopt the action $Y^{(k2)}$
(or, kernel ℓ belongs to the *coarser class* defined by Rule $k2$).

Hence, the proposed rule-based classifier must schedule Rules 1 through 6 in sequence.

A computer program, CERES (named after the Roman Goddess of grain and harvests), has been designed to classify rice kernels based on the preceding six rules. The program is written in WATFIV, and is currently in the process of validation by grain experts. Several hundred milled rice kernels have been classified to this point, and the results have been most satisfactory.

EXPERIMENTAL WORK

The methodology for the combination of evidence in the production rule formalism was applied to the problem of classifying long-grain milled rice into the four classes, *wholes*, *second heads*, *screenings*, and *brewers*, described in the preceding sections. The present approach towards grain classification relies on simple rules of discrimination that define prototype kernels from the four classes. These rules are expressed in terms of parameters that pertain to the shape and size of kernels. The experimental work involved the measurement of the parameters of length, width, area, and slenderness for sample rice kernels. Samples of Arkansas-grown, long-grain milled rice were used in the study.

Equipment

A Quantimet 720B image analyzing computer, manufactured by Image Analyzing Computers (IMANCO), Hertfordshire, England, was employed for measuring the parameters pertaining to the shape and size of rice kernels. To increase the accuracy of measurements, this system deliberately abandons conventional television standards in incorporating digitally controlled 720 line scanners, purpose-built to minimize electronic noise. Additionally, an automatic 686 point matrix shading corrector, coupled with the low noise of special slow-speed scanners, enable the detection of 64 gray levels. The instrument is equipped with the following basic modules for operation in the manual mode:

i) Vidicon Scanner (50 mm lens)

ii) Display

iii) System Control

iv) Frame Generator

v) 1D Auto-Detector

vi) MS3 Standard Computer

vii) Light Pen.

All experimental work was performed with the analyzer located at the United States Grain Marketing Research Laboratory, Manhattan, Kansas.

Measurement Procedure

The Quantimet 720B image analyzer is permanently calibrated in *picture points*, and all measurements are made in these units. For calibration in absolute units, it was necessary only to find the linear equivalent at the specimen of a picture point for the optical system being used. This was accomplished by employing a scale calibrated in millimeters, and computing the picture-point-equivalent.

After the calibration procedure, grain samples were placed on a black background, and were imaged on the Display Screen. Next, the imaged kernels were detected using the 1D Auto-Detector. Subsequently, the MS3 Standard Computer was employed to measure the length, width, and area (projected) of the detected kernels. All measurements were performed in the manual setting, and, therefore, the slenderness parameter, defined as the ratio of kernel length to its width, was computed off-line.

CONCLUDING REMARKS

In applications of classical pattern recognition theory, discriminant functions (see, e.g., Andrews, 1972) are often used as the basis for constructing classification algorithms. Discriminant functions have the property that they partition the pattern or feature space into mutually exclusive regions, each region contributing to the domain of a class.

Suppose that S is the pattern space, and it is required to partition this space into n regions, $s_1, s_2, \ldots, s_n$. A discriminant function, $\Phi_i(\hat{z})$, is defined such that for all points (pattern vectors), $\hat{z}$, within the class defined by s_i, we have

$$\Phi_i(\hat{z}) > \Phi_j(\hat{z}), \quad \forall \hat{z} \in s_i,\ i \neq j. \tag{5.22}$$

Thus, within the region s_i, the i-th discriminant function will have the largest value.

The construction and adjustment of suitable functions are important tasks in the design of a discriminant-function-based classifier. Methods involving parametric or non-parametric statistical techniques are often used for developing proper discriminant functions. Sometimes, distribution-free techniques are employed; here, the functional forms of the discriminant functions (i.e., linear, quadratic, etc.) are assumed before-hand.

The rule-based classifier developed in this chapter operates in much the same way as one relying on discriminant functions. Instead of an assumed polynomial functional form, each production rule itself is a *linguistic discriminant function.* The premise is first evaluated, and

the subsequent estimation of the action certainty, $h(Y^{(i)})$, corresponds to the computation of the value of the discriminant function. Thus,

$$h(Y^{(1)}) \equiv \Phi_1(\hat{z}) \quad \text{defines class } s_1 \quad \text{(wholes)} \tag{5.23a}$$

$$h(Y^{(2)}) \equiv \Phi_2(\hat{z}) \quad \text{defines class } s_2 \quad \text{(second heads)} \tag{5.23b}$$

$$h(Y^{(3)}) \equiv \Phi_3(\hat{z}) \quad \text{defines class } s_3 \quad \text{(screenings)} \tag{5.23c}$$

$$h(Y^{(4)}) \equiv \Phi_4(\hat{z}) \quad \text{defines class } s_4 \quad \text{(brewers)} \tag{5.23d}$$

where

$$\hat{z} = (z_1, z_2, z_3, z_4) \tag{5.23e}$$

is the pattern vector containing the shape parameters of a milled rice kernel, and $h(Y^{(i)})$ is the action certainty for the i-th rule of discrimination.

The present methodology offers advantages that are not available when conventional discriminant function formulations are employed. Perhaps, the most significant advantage is in the flexibility offered by the production rule formalism. Human experts find it comfortable to express their knowledge in the form of linguistic rules. Instead of abstract mathematical functions, the propositions in a production rule are easy to understand, and focus directly on the problem of discrimination. Additionally, the evaluation of premises of the production rules employing the Sugeno Integral permits discrimination and classification in an instance in which the pattern vector, $\hat{z}$, is incomplete. In this situation, the computation of a polynomial discriminant function in classical pattern recognition theory, and the evaluation employing the minimum operator in rule-based formulations, are both meaningless. All things considered,

the present methodology provides a powerful framework for a novel scheme of discriminant-function-based classification.

REFERENCES

1. Adair, C. R., "Production and Utilization of Rice", in Houston, D. F., (Ed.), "Rice: Chemistry and Technology", American Assoc. of Cereal Chemists, Inc., St. Paul, Minnesota (1972).

2. Andrews, H. C., "Introduction to Mathematical Techniques in Pattern Recognition", Wiley-Interscience, New York (1972).

3. Barr, A, and Feigenbaum, E. A., "The Handbook of Artificial Intelligence: Vol. 1", William Kaufman Inc., Los Altos, Calif. (1981).

4. Chaudhuri, U. N., Personal communication (1983).

5. Kulkarni, R. G., Personal communication (1983).

6. Mathewson, P. R., Personal communication (1983).

7. McDermott, J., "R1, a rule-based configurer of computer systems", Artif. Intell., 19, 38-88 (1982).

8. Rangnekar, P. D., Personal communication (1983).

9. Shortliffe, E. H., "Computer-based Medical Computations: MYCIN", American Elsevier, New York (1976).

10. Sugeno, M., "Theory of Fuzzy Integrals and its Applications", Ph.D. Thesis, Tokyo Institute of Technology, Tokyo (1974).

11. U. S. Department of Agriculture, "Rice Situation", RS-13, Economic Research Service (March 1969).

12. U. S. Department of Agriculture, "Rice in the United States: Varieties and Production", Agriculture Handbook No. 289, Agriculture Research Service (June 1973).

13. U. S. Department of Agriculture, "United States Standards for Rice", Federal Grain Inspection Service (July 1983).

14. Ward, A. B., Personal communication (1983).

15. Webb, B. D., and Stermer, R. A., "Criteria of Rice Quality", in Houston, D. F., (Ed.), "Rice: Chemistry and Technology", American Assoc. of Cereal Chemists, Inc., St. Paul, Minnesota (1972).

16. Witte, G. C., "Conventional Rice Milling in the United States", in Houston, D. F., (Ed.), "Rice: Chemistry and Technology", American Assoc. of Cereal Chemists, Inc., St. Paul, Minnesota (1972).

17. Zadeh, L. A., "The concept of a linguistic variable and its application to approximate reasoning, Part I", Inf. Sc., 8, 199-249 (1975a).

18. Zadeh, L. A., "The concept of a linguistic variable and its application to approximate reasoning, Part II", Inf. Sc., 8, 301-357 (1975b).

19. Zadeh, L. A., "The concept of a linguistic variable and its application to approximate reasoning, Part III", Inf. Sc., 9, 43-80 (1975c).

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

The present approach employs elements of monotonic measure theory in modeling the subjective combination of evidence in the production rule formalism. The approach has been motivated by the current emphasis on knowledge-based expert systems and the corresponding field of Knowledge Engineering.

Knowledge is power, and Knowledge Engineering is the technology that promises to make knowledge a valuable commodity. In this chapter, we examine the advantages offered by the present methodology in the light of Knowledge Engineering; and proceed to list our recommendations for future work.

CONCLUSIONS

The knowledge base of an expert system is a repository of human knowledge, and the inference engine - the mechanism that manipulates this knowledge - represents an attempt to mirror the processes involved in human reasoning. Human beings are characterized by their ability to reason in subjective terms. This subjectivity is not only an intrinsic feature of human knowledge, but also enters into the reasoning process. Any attempt at modeling human subjectivity, therefore, involves looking toward ways of representing subjective knowledge, as well as procedures for using this knowledge effectively in reasoning mechanisms.

The principle of monotonicity underlies most human evaluative strategies; and the monotonic measures, employed in this work, offer a convenient means of representing subjective information about the relative weights of propositions in premises of production rules. The measures correspond to the deeper or meta-level knowledge pertaining to the interactions of propositions. They are *a priori* epistemic notions that come into play when the evidence provided by individual, or groups of propositions is weighed and balanced. In many areas of human activity, this kind of information forms a large chunk of human expertise. The proper representation of these concepts, therefore, is of considerable importance from the point of view of Knowledge Engineering.

Having represented the subjective knowledge, the next step in developing a methodology that mirrors human subjective reasoning is to get the inference engine to process this knowledge. The mechanism

that processes the deeper knowledge may operate non-linearly, but must also possess reasonable intuitive justification. The Sugeno Integral, a functional defined on monotonic measure space, has been shown to have these properties. Employed to evaluate premises comprised of AND-connected propositions, it unites the lower-level knowledge pertaining to the truth values of propositions with the meta-level information concerning the relative weights of propositions. The result is a subjectively weighted premise evaluation that is also monotonic.

The monotonicity of the measures and the Sugeno Integral have been shown to provide a viable framework for the representation and treatment of ignorance, and the conservatism that is seen in human evaluations made in its presence. Human beings demonstrate an ability to reason in the presence of incomplete information, and the present methodology introduces this ability into systems relying on a pattern of rule-directed inference.

It might be argued that the Sugeno Integral is just one more functional that has been proposed for evaluating premises of production rules. However, it has one significant property; the Sugeno Integral is an extension of the minimum operator that is conventionally used to evaluate premises consisting of AND-connected propositions. It is often felt that the minimum operator is too conservative, since the evaluation it offers is only as good as the worst attribute indicates. The fact that the evaluation employing the minimum operator is the lower bound for the Sugeno Integral and is equal to the Sugeno evaluation in the absence of meta-level knowledge about the relative weights of propositions (vacuous

belief conditions), is in accordance with the conservatism that is the hallmark of human evaluations made in the presence of ignorance.

The methodology is easily extended to admit multilevel reasoning, and good results have been obtained in its application in a rule-based classifier for long-grain milled rice.

In their pioneering paper on the knowledge-based approach in AI, Minsky and Papert (1974) envision "... progress as coming from better ways to express, recognize, and use diverse and particular forms of knowledge...". The subjectivity that enters in the human combination of evidence may be considered to be a specific local feature of knowledge-organizing-knowledge. The present approach appears to provide an excellent framework for expressing, recognizing, and using this knowledge.

RECOMMENDATIONS

In this section, we outline our recommendations for future work; they are listed under two categories. The first concerns potential areas of application of the methodology developed in this thesis, while the second includes extensions and refinements of the methodology.

Production rules are natural to human understanding of problem solving and decision making, and offer a modular scheme for representing and using domain-specific knowledge. Monotonic measures and the Sugeno Integral give production systems the power to express and manipulate human subjective knowledge. Hence, in addition to areas that are attractive for production rule representations, the methodology is especially useful for application in areas in which subjective decision making is an important feature of human expertise.

The Sugeno Integral has been employed in the present work to weigh and balance the evidence provided by individual propositions. In doing so, it essentially performs a rational and systematic trade-off, and the resulting evaluation is balanced and well-rounded. This property could be used in developing novel lexicographic optimization techniques. The use of production rule formulations for this purpose also has its own benefits.

The design and synthesis of chemical processes is more an art than a science. A considerable portion of expertise in this area consists of heuristics, rules-of-thumb, and empirical associations

that are acquired through extensive experience; this is valuable knowledge that can be captured and preserved effectively by production rules. Good engineering judgment has an important role in design and synthesis. This judgment is characterized by its subjectivity, and is central in making trade-off decisions.

Automatic control of chemical processes is another area in which the present methodology can be implemented successfully. Many chemical systems are too complex to model mathematically. Yet, expert human operators are capable of performing control actions very satisfactorily. Heuristic knowledge guides the operator, and often, control actions are performed although all the feedback information may not be provided. It appears to be a good idea to incorporate operator-specified, linguistic rules as a means for effective control. The ability of the present methodology to provide evaluations in the absence of complete information is significant in such applications.

There are several extensions and refinements that could increase the power and applicability of the methodology for the combination of evidence. In this work, we have been concerned mainly with the tasks of expressing and using subjective knowledge. The acquisition of knowledge is an important aspect of Knowledge Engineering; and future work must focus in this direction.

A learning system seems to be the best way in which the machine could obtain accurate values for measures of groups of propositions. Tazaki (1983), and Yasuhara (1983) have investigated the inverse problem, in which knowledge of the value of the Sugeno Integral is

employed in estimating the values of the monotonic measures. This requires an understanding of the functional behaviour of the Sugeno Integral. Another method involves the use of the conditional fuzzy measure (Sugeno, 1974). This monotonic measure is similar to Bayes' conditional probability (additive monotonic measure) that is often used in constructing conventional learning systems (see, e.g., Fu, 1970).

The use of the present methodology in an expert system requires an explanation facility designed specifically for the evaluation employing the Sugeno Integral. The construction of such a facility needs knowledge of the functional behaviour, and the operations involved in integration.

It seems plausible to employ a measure-theoretic approach to express subjective meta-level knowledge. However, as stated previously, the Sugeno Integral is just one functional defined on monotonic measure space that is employed for our purposes because it appears to mirror certain human evaluative strategies. It is, perhaps, possible to define other functionals having better properties. Detailed psychological investigations of human evaluative behaviour are required so that appropriate functionals can be developed.

REFERENCES

1. Fu, K. S., "Statistical Pattern Recognition", in Mendel, J. M., and Fu, K. S., (Eds.), "Adaptive, Learning, and Pattern Recognition Systems: Theory and Applications", Academic Press, New York (1970).

2. Minsky, M., and Papert, S., "Artificial Intelligence", Condon Lectures, Oregon State System for Higher Education, Eugene, Oregon (1974).

3. Sugeno, M., "Theory of Fuzzy Integrals and its Applications", Ph.D. Thesis, Tokyo Institute of Technology, Tokyo (1974).

4. Tazaki, E., Personal communication (1983).

5. Yasuhara, M., "On Fuzzy-Measure Learning Identification Algorithm", M.S. Thesis, Tokyo University of Science, Tokyo (1983).

APPENDIX I

UNITED STATES STANDARDS FOR MILLED RICE

U.S. STANDARDS FOR MILLED RICE 1/

TERMS DEFINED

§ 68.301 Definition of milled rice.

Whole or broken kernels of rice (*Oryza sative L.*) from which the hulls and at least the outer bran layers have been removed and which contain not more than 10.0 percent of seeds, paddy kernels, or foreign material, either singly or combined.

§ 68.302 Definition of other terms.

For the purposes of these standards, the following terms shall have the meanings stated below:

(a) Broken kernels. Kernels of rice which are less than three-fourths of whole kernels.

(b) Brown rice. Whole or broken kernels of rice from which the hulls have been removed.

(c) Chalky kernels. Whole or broken kernels of rice which are one-half or more chalky.

(d) Classes. There are seven classes of milled rice. The following four classes shall be based on the percentage of whole kernels, (*broken kernels*), and types of rice:

Long-Grain Milled Rice

Medium-Grain Milled Rice

1/ Compliance with the provisions of these standards does not excuse failure to comply with the provisions of the Federal Food, Drug, and Cosmetic Act, or other Federal laws.

Short-Grain Milled Rice

Mixed Milled Rice

The following three classes shall be based on the percentage of whole kernels and of broken kernels of different size:

Second-Head Milled Rice

Screenings Milled Rice

Brewers Milled Rice

(1) "Long-grain milled rice" shall consist of milled rice which contains more than 25.0 percent of whole kernels of milled rice and in U.S. Nos. 1 through 4 not more than 10.0 percent of whole or broken kernels of medium- or short-grain rice. U.S. No. 5 and U.S. No. 6 long-grain milled rice shall contain not more than 10.0 percent of whole kernels of medium- or short-grain milled rice (*broken kernels do not apply*).

(2) "Medium-grain milled rice" shall consist of milled rice which contains more than 25.0 percent of whole kernels of milled rice and in U.S. Nos. 1 through 4 not more than 10.0 percent of whole or broken kernels of long-grain rice or whole kernels of short-grain rice. U.S. No. 5 and U.S. No. 6 medium-grain milled rice shall contain not more than 10.0 percent of whole kernels of long- or short-grain milled rice (*broken kernels do not apply*).

(3) "Short-grain milled rice" shall consist of milled rice which contains more than 25.0 percent of whole kernels of milled rice and in U.S. Nos. 1 through 4 not more than 10.0 percent of whole or broken kernels of long-grain rice or whole kernels of medium-grain rice. U.S. No. 5 and U.S. No. 6 short-grain milled rice shall contain not more

than 10.0 percent of whole kernels of long- or medium-grain milled rice (*broken kernels do not apply*).

(4) "Mixed milled rice" shall consist of milled rice which contains more than 25.0 percent of whole kernels of milled rice and more than 10.0 percent of "other types" as defined in paragraph (i) of this section. U.S. No. 5 and U.S. No. 6 mixed milled rice shall contain more than 10.0 percent of whole kernels of "other types"(*broken kernels do not apply*).

(5) "Second-head milled rice" shall consist of milled rice which, when determined in accordance with §§ 68.303 and 68.304 contains:

(i) Not more than (a) 25.0 percent of whole kernels, (b) 7.0 percent of broken kernels removed by a 6 plate, (c) 0.4 percent of broken kernels removed by a 5 plate, and (d) 0.05 percent of broken kernels passing through a 4 sieve (*southern production*); or

(ii) Not more than (a) 25.0 percent of whole kernels, (b) 50.0 percent of broken kernels passing through a 6½ sieve, and (c) 10.0 percent of broken kernels passing through a 6 sieve (*western production*).

(6) "Screenings milled rice" shall consist of milled rice which, when determined in accordance with §§ 68.303 and 68.304, contains:

(i) Not more than (a) 25.0 percent of whole kernels, (b) 10.0 percent of broken kernels removed by a 5 plate, and (c) 0.2 percent of broken kernels passing through a 4 sieve (*southern production*); or

(ii) Not more than (a) 25.0 percent of whole kernels (b) 15.0 percent of broken kernels passing through a 5½ sieve; and more than (c) 50.0 percent of broken kernels passing through a 6½ sieve and

(d) 10.0 percent of broken kernels passing through a 6 sieve (*western production*).

(7) "Brewers milled rice" shall consist of milled rice which, when determined in accordance with §§ 68.303 and 68.304, contains not more than 25.0 percent of whole kernels and which does not meet the kernel-size requirements for the class Second Head Milled Rice or Screenings Milled Rice.

(e) Damaged kernels. Whole or broken kernels of rice which are distinctly discolored or damaged by water, insects, heat or any other means, and parboiled kernels in nonparboiled rice. "Heat-damaged kernels" [*see paragraph (g) of this section*] shall not function as damaged kernels.

(f) Foreign material. All matter other than rice and seeds. Hulls germs, and bran which have separated from the kernels of rice shall be considered foreign material.

(g) Heat-damaged kernels. Whole or broken kernels of rice which are materially discolored and damaged as a result of heating and parboiled kernels in nonparboiled rice which are as dark as, or darker in color than, the interpretive line for heat-damaged kernels.

(h) Objectionable seeds. Seeds other than rice, except seeds of *Echinochloa crusgalli (commonly known as barnyard grass, watergrass, and Japanese Millet).*

(i) Other types. (1) Whole kernels of: (i) Long-grain rice in medium- or short-grain rice, (ii) Medium-grain rice in long- or short-grain rice, (iii) Short-grain in long- or medium-grain rice, and (2)

broken kernels of long-grain rice in medium- or short-grain rice and broken kernels of medium- or short-grain rice in long-grain rice, except in U.S. No. 5 and U.S. No. 6 milled rice. In U.S. No. 5 and U.S. No. 6 milled rice, only whole kernels will apply.

NOTE: Broken kernels of medium-grain rice in short-grain rice and broken kernels of short-grain rice in medium-grain rice shall not be considered other types.

(j) Paddy kernels. Whole or broken unhulled kernels of rice; whole or broken kernels of brown rice, and whole or broken kernels of milled rice having a portion or portions of the hull remaining which cover one-eighth (1/8) or more of the whole or broken kernel.

(k) Red rice. Whole or broken kernels of rice on which there is an appreicable amount of red bran.

(l) Seeds. Whole or broken seeds of any plant other than rice.

(m) Types of rice. There are three types of milled rice as follows:

Long grain

Medium grain

Short grain

Types shall be based on the length-width ratio of kernels of rice that are unbroken and the width, thickness, and shape of kernels that are broken as set forth in the Rice Inspection Handbook 2/.

(n) Ungelatinized kernels. Whole or broken kernels of parboiled

---

2/ Publications referenced in these standards will be made available upon request to the Federal Grain Inspection Service, U.S. Department of Agriculture, 1400 Independence Avenue, S.W. Washington, D.C. 20250.

rice with distinct white or chalky areas due to incomplete gelatinization of the starch.

(o) Well-milled kernels. Whole or broken kernels of rice from which the hulls and practically all of the germs and bran layers have been removed.

NOTE: This factor is determined on an individual kernel basis and applies to the special grade Undermilled milled rice only.

(p) Whole kernels. Unbroken kernels of rice and broken kernels of rice which are at least three-fourths of an unbroken kernel.

(q) 5 plate. A laminated metal plate 0.142-inch thick, with a top lamina, 0.051-inch thick, perforated with rows of round holes 0.0781 (5/64) inch in diameter, 5/32 inch from center to center, with each row staggered in relation to the adjacent rows, and a bottom lamina 0.091-inch thick, without perforations.

(r) 6 plate. A laminated metal plate 0.142-inch thick, with a top lamina, 0.051-inch thick, perforated with rows of round holes 0.0938 (6/64) inch in diameter, 5/32 inch from center to center, with each row staggered in relation to the adjacent rows, and a bottom lamina 0.091-inch thick, without perforations.

(s) 2½ sieve.A metal sieve 0.032-inch thick, perforated with rows of round holes 0.0391 (2½/64) inch in diameter, 0.075-inch from center to center, with each row staggered in relation to the adjacent rows.

(t) 4 sieve. A metal sieve 0.032-inch thick, perforated with rows of round holes 0.0625 (4/64) inch in diameter, 1/8 inch from center to center, with each row staggered in relation to the adjacent rows.

(u) 5 sieve. A metal sieve 0.032-inch thick, perforated with rows of round holes 0.0781 (*5/64*) inch in diameter, 5/32 inch from center to center, with each row staggered in relation to the adjacent rows.

(v) 5½ sieve. A metal sieve 0.032-inch thick, perforated with rows of round holes 0.0859 (*5½/64*) inch in diameter, 9/64 inch from center to center, with each row staggered in relation to the adjacent rows.

(w) 6 sieve. A metal sieve 0.032-inch thick, perforated with rows of round holes 0.0938 (*6/64*) inch in diameter, 5/32 inch from center to center, with each row staggered in relation to the adjacent rows.

(x) 6½ sieve. A metal sieve 0.032-inch thick, perforated with rows of round holes 0.1016 (*6½/64*) inch in diameter, 5/32 inch from center to center with each row staggered in relation to the adjacent rows.

(y) 30 sieve. A woven wire cloth sieve having 0.0234-inch openings, with a wire diameter of 0.0153-inch, and meeting the specifications of American Society for Testing and Materials Designation E-11-61, as set forth in the Equipment Handbook 2/.

PRINCIPLES GOVERNING APPLICATION OF STANDARDS

§ 68.303 Basis of Determination.

All determinations shall be on the basis of the original sample. Mechanical sizing of kernels shall be adjusted by handpicking, as set forth in the Rice Inspection Handbook 2/, or by any other method which gives equivalent results.

§ 68.304 Temporary modifications in equipment and procedures.

The equipment and procedures referenced to in the milled rice standards

are applicable to rice produced and harvested under normal environmental conditions. Abnormal environmental conditions during the production and harvest of rice may require minor temporary modifications in the equipment or procedures to obtain results expected under normal conditions. When these adjustments are necessary, Federal Grain Inspection Service Field Offices, official inspection agencies, and interested parties in the rice industry will be notified promptly in writing of the modification. These modifications shall not include changes in interpretations of identity, class, quality, or condition.

§ 68.305 Broken kernels determination.

Broken kernels shall be determined by the use of equipment and procedures set forth in the Rice Inspectation Handbook 2/, or by any method which gives equivalent results.

§ 68.306 Interpretive line samples.

Interpretive line samples showing the official scoring line for factors that are determined by visual observation shall be maintained by the Federal Grain Inspection Service, U.S. Department of Agriculture, and shall be available for reference in all inspections offices that inspect and grade rice.

§ 68.307 Milling requirements.

The degree of milling for milled rice; i.e., "well milled," "reasonably well milled," and "lightly milled" shall be equal to, or better than that of the interpretive line samples for such rice.

§ 68.308 Moisture.

Water content in milled rice as determined by an approved device in

accordance with procedures prescribed in the Rice Inspection Handbook 2/. For the purpose of this paragraph, "approved device" shall include the Motomco Moisture Meter and any other equipment that is approved by the administrator as giving equivalent results 3/.

§ 68.309 Percentages.

Percentages shall be determined on the basis of weight and shall be rounded off as follows:

(a) When the figure to be rounded is followed by a figure greater than 5, round to the next higher figure; e.g., 0.46, report as 0.5.

(b) When the figure to be rounded is followed by a figure less than 5, round to the next lowest figure; e.g., 0.54, report as 0.5.

(c) When the figure to be rounded is followed by the figure 5, round to the nearest even figure; e.g., 0.45, report as 0.4; 0.55, report as 0.6.

All percentages, except for milling yield, shall be stated in whole and tenth percent to the nearest tenth percent. Milling yield shall be stated to the nearest whole percent.

GRADES, GRADE REQUIREMENTS, AND GRADE DESIGNATIONS

For §§§§ 68.310, 68.311, 68.312, and 68.313 see United States Standards for Rice, revised July 1983, pp 23-26, U.S. Department of Agriculture,

---

3/ Requests for information concerning approved devices and procedures, criteria for approved devices, and requests for approval of devices should be directed to the Federal Grain Inspection Service, U.S. Department of Agriculture, 1400 Independence Avenue, S.W., Washington, D.C. 20250.

Federal Grain Inspection Service, Washington, D.C..

§ 68.314 Grade Designations.

(a) The grade designation for all classes of milled rice, except Mixed Milled Rice, shall include in the following order: (1) The letters "U.S."; (2) the number of the grade or the words "Sample grade", as warranted; (3) the words "or better", when applicable and requested by the applicant prior to inspection; (4) the class; and (5) each applicable special grade (*see* § *68.316*).

(b) The grade designation for the class Mixed Milled Rice shall include, in the following order: (1) The letters "U.S."; (2) the number of the grade or the words "Sample grade," as warranted; (3) the words "or better," when applicable and requested by the applicant prior to inspection; (4) the class; (5) each applicable special grade (*see* § *68.316*); (6) the percentage of whole kernels of each type in the order of predominance and when applicable; (7) the percentage of broken kernels of each type in the order of predominance; and (8) the percentage of seeds and foreign material.

NOTE: Broken kernels other than long grain, in Mixed Milled Rice, shall be certificated as "medium or short grain".

SPECIAL GRADES, SPECIAL GRADE REQUIREMENTS, SPECIAL GRADE DESIGNATIONS

§ 68.315 Special grade and special grade requirements.

A special grade when applicable, is supplemental to the grade assigned under § 68.314. Such special grades for milled rice are established and determined as follows:

(a) Coated milled rice. Coated milled rice shall be rice which is coated, in whole or in part, with substances that are safe and suitable 4/ according to commercially accepted practice.

(b) Granulated brewers milled rice. Granulated brewers milled rice shall be milled rice which has been crushed or granulated so that 95.0 percent or more will pass through a 5 sieve, 70.0 percent or more will pass through a 4 sieve, and not more than 15.0 percent will pass through a 2½ sieve.

(c) Parboiled milled rice. Parboiled milled rice shall be milled rice in which the starch has been gelatinized by soaking, steaming, and drying. Grades U.S. No. 1 to U.S. No. 6, inclusive, shall contain not more than 10.0 percent of ungelatinized kernels. Grades U.S. No. 1 and U.S. No. 2 shall contain not more than 0.1 percent, grades U.S. No. 3 and U.S. No. 4 not more than 0.2 percent, and grades U.S. No. 5 and U.S. No. 6 not more than 0.5 percent of nonparboiled rice. If the rice is: (1) Not distinctly colored by the parboiling process, it shall be considered "Parboiled Light"; (2) distinctly but not materially colored by the parboiled process, it shall be considered "Parboiled"; (3) materially colored by the parboiled process, it shall be considered "Parboiled Dark". The color levels for "Parboiled Light", "Parboiled", and "Parboiled Dark" shall be in accordance with the interpretive line samples for parboiled

---

4/ Compliance with the provisions of these standards does not excuse failure to comply with provisions of the Federal Food, Drug, and Cosmetic Act, or other Federal Laws. Safe and suitable is defined in the regulation issued pursuant to the Federal Food, Drug and Cosmetic Act at 21 CFR 130.3(d).

rice.

NOTE: The maximum limits for "Chalky kernels", "Heat-damaged kernels", "Kernels damaged by heat", and the "Color requirements" in §§§§ 68.310, 68.311, 68.312, and 68.313 are not applicable to the special grade "Parboiled milled rice".

(d) Undermilled milled rice. Undermilled milled rice shall be milled rice which is not equal to the milling requirements for "well milled", "reasonably well milled", and "lightly milled" rice (*see § 68.307*). Grades U.S. No. 1 and U.S. No. 2 shall contain not more than 2.0 percent, grades U.S. No. 3 and U.S. No. 4 not more than 5.0 percent, grade U.S. No. 5 not more than 10.0 percent, and grade U.S. No. 6 not more than 15.0 percent of well-milled kernels. Grade U.S. No. 5 shall contain not more than 10.0 percent of red rice and damaged kernels (*singly or combined*) and in no case more than 6.0 percent of damaged kernels.

NOTE: The "Color and milling requirements" in §§§§ 68.310, 68.311, 68.312, and 68.313 are not applicable to the special grade "Undermilled milled rice".

§ 68.316 Special Grade designation.

The grade designation for coated, granulated brewers, parboiled, or undermilled milled rice shall include, following the class, the word(s) "Coated", "Granulated", "Parboiled Light", "Parboiled", "Parboiled Dark", or "Undermilled", as warranted, and all other information as prescribed in § 68.314.

[These standards are taken from "United States Standards for Rice", revised July 1983, U.S. Department of Agriculture, Federal Grain Inspection Service, Washington, D.C..]

MONOTONIC MEASURES IN KNOWLEDGE ENGINEERING:
AN APPLICATION IN THE PRODUCTION RULE FORMALISM

by

SUJEET SHENOI

B.Tech., Indian Institute of Technology, Bombay, 1981

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Chemical Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1984

ABSTRACT

The goal of research in Artificial Intelligence (AI) is to get machines to emulate intelligent human behaviour. Knowledge Engineering, a sub-field of AI, focuses on the technical issues of acquiring, representing, and using knowledge in constructing computer programs that can "reason". This has led to the development of expert systems. Modeled on human experts, these programs embody knowledge, and use it to solve real-world problems in specific areas of human activity. Production rules (or IF-THEN rules) are a popular approach for representing and manipulating domain knowledge in expert systems. Implemented in rule-based or production systems, they are natural to human problem solving strategies.

Human beings are able to reason subjectively, and this ability must be incorporated in any synthetic model of human reasoning. In making evaluations, humans tend to weigh and balance the evidence they receive, and this feature may be assumed to introduce subjectivity. The present work focuses on the combination of evidence in the production rule formalism. The premise of a rule, comprised of AND-connected propositions, is written as a set, and each proposition is a distinct piece of evidence pointing to the action. Some propositions are more important than others, and monotonic measures are employed to hold meta-level information pertaining to the subjective weights of propositions. The evaluation of the premise, therefore, requires the combination of truth values of individual propositions with their

relative weights. The Sugeno Integral, a functional defined on a monotonic measure space, unites these two quantities. The result is a subjectively weighted premise evaluation that is also monotonic, and has excellent intuitive justification.

The introduction of monotonic measure theory into the production rule formalism provides a logical foundation for expressing and coping with the subjectivity that is the hallmark of human evaluative strategies. It offers a viable framework for the representation and treatment of ignorance. Additionally, the Sugeno Integral is simply an extension of the conventional minimum operator, and the methodology can be extended to admit multilevel reasoning. The development of the present methodology is in keeping with the guidelines of Knowledge Engineering, and the advantages gain significance when viewed in this light.

The methodology is implemented in a rule-based system for the classification of long-grain milled rice. Several hundred unclassified kernels have been evaluated, and satisfactory results have been obtained.